John von Neumann Institute for Computing

Tensors in Electronic Structure Theory: Basic Concepts and Applications to Electron Correlation Models

Martin Head-Gordon, Michael Lee, Paul Maslen,
Troy van Voorhis, and Steven Gwaltney

published in

Modern Methods and Algorithms of Quantum Chemistry,
Proceedings, Second Edition, J. Grotendorst (Ed.),
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **3**, ISBN 3-00-005834-6, pp. 593-638, 2000.

http://www.fz-juelich.de/nic-series/

TENSORS IN ELECTRONIC STRUCTURE THEORY: BASIC CONCEPTS AND APPLICATIONS TO ELECTRON CORRELATION MODELS

MARTIN HEAD-GORDON, MICHAEL S. LEE, PAUL E. MASLEN,
TROY VAN VOORHIS, AND STEVEN R. GWALTNEY

Department of Chemistry, University of California, and
Chemical Sciences Division, Lawrence Berkeley National Laboratory,
Berkeley, CA 94720, USA
E-mail: mhg@bastille.cchem.berkeley.edu, mslee@scripps.edu
{maslen,tvan,gwaltney}@bastille.cchem.berkeley.edu

Electronic structure theories correspond to approximate Schrödinger equations, which are usually solved within finite basis expansions. Most textbook treatments of electronic structure theory present both derivations and working equations in terms of a single particle basis of orthogonal functions. All that matters, however, is the space that the functions collectively span, and observables such as the energy are naturally invariant to quite general nonunitary transformations of the underlying basis functions. Tensor methods compactly express such invariances. This chapter presents a simple introduction to tensor methods and their applications in electronic structure theory. One important message is that there is no extra algebraic effort necessary to derive electronic structure theories in terms of an underlying non-orthogonal basis when the so-called natural representation is used. In this representation there is a term-by-term correspondence with equations in an orthogonal basis. It is then straightforward to transform such equations via metric matrices into the covariant integral representation, which is sometimes more convenient for computational purposes, or as a starting point for further approximations, such as local correlation models. In the remainder of the chapter, the development of local electron correlation methods using nonorthogonal functions to span both the occupied and virtual spaces is discussed as an application of the tensor methods, and as an interesting new methodology in its own right. Additionally several other recent uses of the tensor methods from our group are briefly summarized.

1 Introduction

In this chapter, we discuss the development of electronic structure theory without requiring any orthogonalization of the underlying one-particle basis, which, in general, is naturally nonorthogonal. We employ tensor methods to permit the treatment of nonorthogonality in an efficient and general fashion. The resulting tensor equations express the working equations of any electronic structure method in a way that exhibits all of the natural invariances of such equations to underlying transformations of the basis, between either different orthogonal or nonorthogonal representations. By contrast, standard textbook treatments[1,2,3] of electronic structure theory tend to present working equations that are explicitly in an orthogonal basis, and perhaps in the end back-transform them to the original atomic orbital basis. The basic reason for employing orthogonal representations is that they appear to yield simpler equations: working equations in the atomic orbital basis are complicated by the appearance of the overlap matrix in many places. Part of the beauty of the tensor-based approach advocated here is that it allows us to obtain ex-

pressions that are general to nonorthogonal basis sets with essentially no additional complication. This has been recognized in the research literature[4,5,6,7], although it is generally not widely known amongst electronic structure theorists.

The usefulness of working equations in the nonorthogonal atomic orbital representation (or perhaps some other local nonorthogonal representation) needs little explanation these days. The atomic orbital basis consists of strongly localized functions, which permits any localization of the one and two-particle density matrices in real-space to be exploited for computational efficiency. For example, recent advances in linear scaling methods[8,9] for mean field electronic structure calculations depend critically on the use of localized nonorthogonal functions. Localized (and generally nonorthogonal) single particle functions also provide a natural representation in which the description of electron correlation via many-body theories can be cast in local terms. A spatially localized treatment of electron correlation[10,11] is one way to reduce the unphysical scaling of computational cost with molecular size that afflicts orthogonal basis formulations of many-body theories.

Beyond pragmatic considerations of computational efficiency, perhaps the primary purpose of allowing nonorthogonal functions is to obtain expressions for electronic structure theories that are as general as possible. By this we mean that an electronic structure theory normally corresponds to solving a set of equations within some given one, two or many-particle Hilbert space: this is the working model of the Schrodinger equation for a given model chemistry. The details of the individual functions that span these spaces are clearly unimportant in general: all that matters is the space they collectively define. In other words, what we are saying is that electronic structure equations are usually operator relations, which are merely represented in a given basis. Tensor methods have the important advantage of naturally expressing this invariance to nonsingular transformations of the basis functions. It is intuitive that the operators and approximate wave functions depend only on the vector space spanned by the basis functions, rather than details of the basis functions such as nonorthogonality. Their matrix representations are the tensors that we shall focus on.

The first part of this chapter is an introduction to simple concepts of tensor analysis, which does not assume any background in the area. In terms of electronic structure methods, this then lets us more or less immediately treat problems that involve functions of only one electron at a time, as an application of standard tensor methods to quantum mechanics in a finite-dimensional one-particle Hilbert space. We emphasize the fact that exact linear dependence in the basis does not pose a significant problem. The next major topic is the treatment of many-body methods as an application of tensor theory. In first quantization, generalizations of Slater's rules for matrix elements are obtained, while in second quantization, Wick's theorem holds, and lead to representations of one and two particle operators in terms of the quasiparticle reference. These two topics comprise the tutorial part of the chapter (Sections 2 and 3).

The next major part of the chapter (Section 4) describes the use of tensor concepts to treat electron correlation in a spatially localized representation. We discuss new methods that reduce the complexity of electron correlation theories by making physically motivated truncations of the wavefunction variables. In partic-

ular, the simplest wavefunction-based description of electron correlation is second order Møller-Plesset (MP2) theory. In MP2 theory, the wavefunction variables are amplitudes that describe correlated fluctuations of pairs of electrons from a mean field reference into excited states. The tensor approach allows us to represent the spaces that are occupied and unoccupied in the mean-field reference in terms of over-complete sets of atom-centered functions. In turn, this means that the double substitution amplitudes can now be represented in terms of these atom-centered functions. If no local truncation is performed, then the MP2 energy would be more expensive to evaluate in this representation than in terms of the usual "canonical" molecular orbitals (at least naively).

The overcomplete atomic representations are chosen because they are an ideal starting point for local truncations of the double substitution amplitudes. We discuss in detail the various local truncations that can be obtained by restricting the number of amplitudes retained, based on an atomic criterion. An example of such a criterion is that at least one occupied and one unoccupied index of a double substitution must belong to the same atom, otherwise that amplitude is set to zero. Such atomic truncations yield simplified "local" descriptions of the electron correlations, with inherently reduced computational complexity. Furthermore, atomic truncations have the important advantage that they yield inherently smooth potential energy surfaces, and, indeed the resulting local correlation methods satisfy all the criteria of a well-defined theoretical model chemistry. It may at first sound like these atomic truncations are very drastic, and so we additionally present a selection of numerical results to show that the resulting "local model chemistries" are quite faithful to the model chemistry obtained without any truncation (within the MP2 model).

There are, of course, innumerable other possible applications of tensor concepts to outstanding problems in electronic structure theory, either of the model development type, or of the algorithmic type. The final part of this chapter (Section 5) is a short introduction to several other areas where we have found these concepts to be helpful. This includes the simplified description of geminal wavefunctions for bond-breaking, extensions of the local correlation models discussed for MP2 theory to triple substitutions, and perturbation theory with non-Hermitian Hamiltonians. The intention is to give the reader some flavor of why nonorthogonal orbitals and tensor concepts are useful in these problems. It is important to emphasize that there are many other interesting and important applications of tensor concepts in electronic structure theory as well. Some have been pursued by other groups already, but many are yet to be investigated at all! For this reason, we believe that a general knowledge of tensor methods is useful for someone who is planning to do research in electronic structure theory, and the purpose of this chapter is to provide a starting point.

2 Basic Tensor Concepts

While we must refer to textbooks[12,13,14,15] for a full introduction to tensor analysis, we shall develop the basic concepts necessary for application to electronic structure methods in this section.

2.1 Covariant Basis Functions and the Metric

The introduction of a finite basis of one-particle functions is the so-called algebraic approximation of electronic structure theory. The one-particle basis functions are atomic orbitals, which are non-orthogonal amongst themselves. Let us term this set of given functions the covariant basis, and write them as: $\{|\phi_\mu\rangle\}$. While it is not necessary, for simplicity we shall assume that the covariant basis functions (and indeed all matrix elements involving the basis functions) are real. Other sets of one-particle covariant functions will also be used later. They will be for example a set of functions that spans only the occupied part of the one-particle space, or only the unoccupied part.

Let us introduce a first basic definition. Quantities which, upon a transformation of the basis, change in the same way as the basis functions, are termed *covariant. Covariant quantities are denoted via subscripts.* All matrix elements involving the basis functions are entirely covariant in character as they are calculated in a quantum chemistry program. In particular, the overlap matrix plays a very central role in tensor analysis and is called the *covariant metric*. It will be denoted as $g_{\mu\nu}$:

$$g_{\mu\nu} \equiv S_{\mu\nu} = \langle \phi_\mu \mid \phi_\nu \rangle \tag{1}$$

2.2 Contravariant Basis Functions and the Inverse Metric

Given a nonorthogonal set of basic functions, or, in tensor language, given a covariant basis, how will we be able to resolve a vector into components? In other words, how will the operation of projection be performed in the absence of orthogonality? The answer is that there is a matching "dual" basis that can be readily derived from the covariant functions, whose members have the property of being biorthogonal to the covariant functions.

These functions, which are called contravariant basis functions, are defined by the action of the inverse overlap matrix (or the inverse of the covariant metric), on the covariant basis functions:

$$|\phi^\mu\rangle = \sum_\nu |\phi_\nu\rangle \left(S^{-1}\right)^{\nu\mu} \tag{2}$$

The term *contravariant* is meant to imply that these functions transform on a change of basis in the opposite or inverse manner to the way in which the covariant functions transform. These transformation properties will be established shortly in the following subsection. To distinguish contravariant functions from covariant, *contravariant indexes are written as superscripts.*

A one-line proof verifies that, by construction, the contravariant functions are indeed biorthogonal to the covariant functions:

$$\langle \phi^\mu \mid \phi_\nu \rangle = \sum_\lambda \left(S^{-1}\right)^{\mu\lambda} \langle \phi_\lambda \mid \phi_\nu \rangle = \delta^\mu_\nu \tag{3}$$

The overlap matrix of the contravariant functions is in fact the inverse of the overlap matrix of the covariant functions:

$$g^{\mu\nu} \equiv \langle \phi^{\mu} \mid \phi^{\nu} \rangle = \sum_{\lambda\sigma} \left(S^{-1}\right)^{\mu\lambda} \langle \phi_{\lambda} \mid \phi_{\sigma} \rangle \left(S^{-1}\right)^{\sigma\nu} = \left(S^{-1}\right)^{\mu\nu} \tag{4}$$

The role of the metric matrices can now be clearly seen. The contravariant metric defined above has the general property that it converts a covariant index to a contravariant one. Specifically we now see that Eq. (2) can be rewritten as:

$$|\phi^{\mu}\rangle = \sum_{\nu} |\phi_{\nu}\rangle g^{\nu\mu} \tag{5}$$

Likewise the covariant metric matrix acting on a contravariant index converts that index to being of the covariant type:

$$|\phi_{\mu}\rangle = \sum_{\nu} |\phi^{\nu}\rangle g_{\nu\mu} \tag{6}$$

as is immediately proven by substituting Eq. (2) into Eq. (6). Thus the *metric matrices allow interconversion of covariant and contravariant indices.*

2.3 Invariances to Transformations and the Summation Convention

Suppose we transform from the original set of covariant basis functions, $\{|\phi_{\mu}\rangle\}$, to a modified set, $\{|\tilde{\phi}_{\mu}\rangle\}$, spanning the same space by a transformation matrix $\mathbf{T}$:

$$|\tilde{\phi}_{\mu}\rangle = \sum_{\nu} |\phi_{\nu}\rangle T^{\nu}_{\bullet\mu} \tag{7}$$

The placeholder ($\bullet$), means that the first index of the transformation is contravariant while the second index is covariant, to distinguish them since they are inequivalent. As discussed further in a subsequent subsection, we implicitly adopt the convention that the right-hand index has the character of a ket and the left-hand index has the character of a bra, in terms of Dirac's bra-ket notation.

This new covariant basis will have an overlap matrix, or covariant metric, which is related to the previous one by the following expression:

$$\tilde{g}_{\mu\nu} = \langle \tilde{\phi}_{\mu} \mid \tilde{\phi}_{\nu} \rangle = \sum_{\lambda\sigma} T^{\bullet\lambda}_{\mu} g_{\lambda\sigma} T^{\sigma}_{\bullet\nu} \tag{8}$$

For real transformations $\mathbf{T}$, the adjoint of $\mathbf{T}$ is the same as the transpose; i.e. $T^{\bullet\nu}_{\mu} = T^{\nu}_{\bullet\mu}$.

Given this new set of covariant basis functions, we can define a new contravariant basis by the prescription given in the previous subsection.

$$|\tilde{\phi}^{\mu}\rangle = \sum_{\nu} |\tilde{\phi}_{\nu}\rangle \tilde{g}^{\nu\mu} \tag{9}$$

How are these new contravariant functions related to the original set? The inverse of the new metric defined by Eq. (8) is evidently related to the original inverse metric by the following relation involving the inverse of the transformation, **T**:

$$\tilde{g}^{\mu\nu} = \sum_{\lambda\sigma} \left(T^{-1}\right)^{\mu}_{\bullet\lambda} g^{\lambda\sigma} \left(T^{-1}\right)^{\bullet\nu}_{\sigma} \tag{10}$$

Substituting Eqs. (8) and (10) into Eq. (9) we obtain the transformation property of the contravariant functions:

$$\left|\tilde{\phi}^{\mu}\right\rangle = \sum_{\nu} \left|\phi^{\nu}\right\rangle \left(T^{-1}\right)^{\bullet\mu}_{\nu} \tag{11}$$

The *contravariant functions transform inversely to the way that the covariant functions transform.* More generally, therefore, covariant and contravariant *indices* have inverse transformation properties.

This fact is important in tensor analysis. If we form a scalar quantity (or more generally reduce the number of free indices by one) by summing over one covariant index and one contravariant index:

$$c = \sum_{\mu} a^{\mu} b_{\mu} \tag{12}$$

then such a scalar will be invariant to transformations of the basis:

$$\tilde{c} = \sum_{\mu} \tilde{a}^{\mu} \tilde{b}_{\mu} = \sum_{\mu\nu\lambda} a^{\nu} \left(T^{-1}\right)^{\bullet\mu}_{\nu} T^{\bullet\lambda}_{\mu} b_{\lambda} = \sum_{\nu\lambda} a^{\nu} \delta^{\bullet\lambda}_{\nu} b_{\lambda} = c \tag{13}$$

The energy in an electronic structure theory is just such a scalar, which will be given by various sums over orbitals. Tensor notation will therefore immediately express the invariance of the energy to certain classes of transformations of the orbitals.

The *summation convention* simply states that the simple presence of a repeated index implies summation, provided *the index occurs once in covariant form and once in contravariant form*, as needed if the result is to be invariant. Specifically, we are defining:

$$a^{\mu} b_{\mu} \equiv \sum_{\mu} a^{\mu} b_{\mu} \tag{14}$$

So, at this stage we have a useful perspective. The introduction of covariant and contravariant indices that have inverse properties upon transformations of the basis permits us to sum over indices in a way that is invariant with respect to such transformations. Such a summation is termed a contraction.

More generally, tensor notation, meaning the use of the covariant and contravariant representations, will let us write express operator relations such that

their transformation properties upon a change of basis are fully implicit in the algebra. Certain rules must be followed for this to be true:

(1) Contraction must only be over index pairs where one index is covariant and one index is contravariant.
(2) Addition of one quantity to another will only be valid if they have indices that match one-to-one in covariant and contravariant character.

2.4 Flat Euclidean Spaces Versus Curved Spaces

A brief word on the geometrical character of what we have discussed to this stage is in order. We have introduced a basis, with a metric or overlap matrix that is nondiagonal. This corresponds to a set of axes that are not orthogonal to each other, but are instead skewed. We are treating the elements of the metric as constants whose value is the same regardless of where we are in the single particle space. Thus we are working in Euclidean geometry, and space is flat. For simplicity we shall retain this restriction throughout this article.

Other choices of basis vectors may define a metric whose values change depending on where we are in the space. The basis vectors corresponding to spherical polar coordinates in 3-space are a familiar example. Another example that is familiar in chemistry are the internal coordinates often used to describe molecular geometry. These basis vectors define spaces that are curved rather than flat, and it is important to properly account for this curvature. For example, to describe the shortest distance between two points, straight lines must be replaced by geodesics (the generalization of great circles). Additionally vectors can no longer be simply translated through space to change their origin, but instead must change their orientation with translation. For further consideration of this topic, we refer the reader to introductory (or not so introductory!) treatments of differential geometry[12,14,15]. An interesting and detailed recent article[16] has much relevance to minimizing self-consistent field type energy expressions.

2.5 Bras, Kets and Operators as Tensors

When we form matrix elements in electronic structure theory, we are, by definition, combining bras and kets. Therefore the tensors which we shall employ to represent operators will have associated with them a set of indices that will be divided into equal numbers of bra indices and ket indices. We adopt the useful convention that for a matrix element involving $2n$ indices, the first n indices will be associated with bras and the last n indices will be associated with kets. That is consistent with how all the matrix elements have been written in the previous parts of this section.

Let us consider the representation of some simple operators. Most basic is the operator for projection onto the space spanned by the covariant basis. The form of this projection operator follows from the biorthogonality of the covariant and contravariant functions:

$$\hat{1} = |\phi_\mu\rangle \langle\phi^\mu| \tag{15}$$

The metric matrices may be used to write this relation entirely in terms of either just the covariant basis functions or just the contravariant basis functions:

$$\hat{1} = |\phi_\mu\rangle\, g^{\mu\nu}\, \langle\phi_\nu| = |\phi^\mu\rangle\, g_{\mu\nu}\, \langle\phi^\nu| \tag{16}$$

It is simply verified that any of the three equivalent forms for the projection operator acting on a ket in the covariant space gives back that same ket. They express the "resolution of the identity" in this space. By simply inserting Eq. (15) or Eq. (16) after a bra or before a ket, one resolves it into components in either the covariant or contravariant basis in the usual way.

Expressions for other operators may also be expressed in terms of the covariant and contravariant basis functions by employing the resolution of the identity in any of the three forms given above. For example, using the simplest form, Eq. (15), to resolve a one electron operator into this nonorthogonal basis set yields:

$$\hat{F}\left(\mathbf{r_1}, \mathbf{r'_1}\right) = |\phi_\mu\rangle\, \langle\phi^\mu|\, \hat{F}\, |\phi_\nu\rangle\, \langle\phi^\nu| = |\phi_\mu\rangle\, F^\mu_{\bullet\nu}\, \langle\phi^\nu| \tag{17}$$

In the second (more compact) form of this equation, we are applying the convention that in a tensor of rank 2, the first index corresponds to the bra and the second index corresponds to the ket. A two electron operator may be treated similarly, making sure to apply the resolution of the identity to both the first and second electronic coordinates:

$$\hat{G} = |\phi_\mu\,(1)\rangle\, |\phi_\nu\,(2)\rangle\, \langle\phi^\mu\phi^\nu|\, \phi_\lambda\phi_\sigma\rangle\, \langle\phi^\sigma\,(2)|\, \left\langle\phi^\lambda\,(1)\right| \tag{18}$$

2.6 Natural Representation and the Covariant Integral Representation

Given the fact that a given ket can be represented in either the covariant or contravariant basis, as can a given bra, there are indeed many ways that a given tensor can be written. Specifically a one particle operator, such as given in Eq. (17), can be represented in 2^2=4 ways, while a two-particle operator, such as Eq. (18), can be represented in 2^4=16 different ways. Two of these ways are most useful in the context of electronic structure theory.

The first useful representation is called the *natural representation*, and adopts the convention that all ket indices are represented in the covariant basis, while all bra indices are represented in the contravariant basis. This representation is faithful in the sense that the tensor algebra representing an operator equation will be isomorphic to the operator form of the equation. Crucially, this representation also yields equations that correspond term-by-term to the form of the equations in an orthogonal basis, with the only difference lying in the covariant and contravariant character of individual indices. It is most convenient to obtain tensor equations initially in the natural representation, as they can sometimes be written down by inspection, or inferred from known equations in the orthogonal representation. The essential point is that obtaining the defining equations of an electronic structure method in terms of nonorthogonal functions in the natural representation is no harder than obtaining the equations in a conventional orthogonal representation!

We shall further develop the analogies between the natural representation and conventional orthogonal basis tools for second quantization and many-electron theory in the following sections.

The second useful representation may be called the covariant integral representation, and is the representation in which all matrix elements of one and two particle operators in the Hamiltonian are represented entirely in terms of covariant indices. This is a useful representation because in electronic structure codes, as we have already discussed, these matrix elements are always calculated with all indices covariant. If all matrix elements are defined fully covariant, then the variables that they are contracted with in the equations of a given electronic structure method must be fully contravariant. For example, in one-electron theories, the one-particle density matrix is fully contravariant in this representation. So, to summarize, as a general rule, we will *derive* in the natural representation, but *compute* in the covariant integral representation.

It is worthwhile at this point to briefly contrast what has been done with the tensor notation relative to conventional treatments of nonorthogonality in quantum chemistry. Conventionally what is done is to *assume* a particular tensor representation, such as the covariant integral representation, for the matrix elements and unknowns. This has two principal drawbacks. First, the overlap matrix must be carried explicitly, which makes the equations and derivations cumbersome relative to simply transforming the tensor character of the indices of the final equations in the natural representation with metric matrices. Second, without the explicit covariant and contravariant character of the indices, it is all too easy to violate the rules of tensor algebra.

2.7 Treatment of Exact Linear Dependence

We must extend the tensor treatment to the case where exact linear dependence exists in the single particle basis. The reason for doing so (rather than eliminating such redundancy at the outset) is that often convenient sets of occupied and virtual functions may include exact linear dependence. A simple example are *projected atomic orbitals*, defined by acting with P and Q on the parent set of atomic orbitals:

$$|\phi_i\rangle = \hat{P}\,|\phi_\mu\rangle\,\delta_i^\mu = |\phi_\mu\rangle\,P_{\bullet\nu}^{\mu}\delta_i^\nu \tag{19}$$

$$\big|\phi_a\big\rangle = \hat{Q}\big|\phi_\mu\big\rangle\delta_a^\mu = \big(\hat{I}-\hat{P}\big)\big|\phi_\mu\big\rangle\delta_a^\mu = \big|\phi_\mu\big\rangle\big(\delta_a^\mu - P_{\bullet\nu}^{\mu}\delta_a^\nu\big) \tag{20}$$

The delta functions merely preserve index conventions. Projected atomic functions are localized to the same extent as the density matrix itself[17]. Their covariant metric matrices are the covariant representations of P and Q respectively:

$$g_{ij} = \langle\phi_i \mid \phi_j\rangle = \delta_i^\mu P_{\mu\nu}\delta_j^\nu \tag{21}$$

$$g_{ab} = \langle\phi_a \mid \phi_b\rangle = \delta_a^\mu\,(g_{\mu\nu} - P_{\mu\nu})\,\delta_b^\nu \tag{22}$$

Contravariant functions may be defined by projection on the contravariant basis vectors with P and Q; their contravariant metrics are the contravariant representations of P and Q respectively.

The linear dependence associated with these representations is immediately evident. Take for example the occupied space. The *dimension* (number of linearly independent functions) of the occupied subspace is the particle number, n, but Equations (19) yields a larger set of functions (N, equal to the dimension of the atomic orbital basis itself). The metric matrix, Equation (21), is then not formally invertible, and the ability to change indices from covariant to contravariant seems to be lost. A coordinate transformation from a set of nonredundant functions to a linearly dependent representation (or vice versa) involves a rectangular matrix, which also has no formal inverse. Can we allow these redundant representations in the tensor formalism?

The answer is yes, for the following reason[18]. Trial vectors on which the inverse operates lie in the *range* of the metric (the linearly independent subspace) with no component in the *nullspace*, since a linearly dependent representation still spans the same space as a related linearly independent basis. Hence we employ a *generalized inverse* based on discarding the nullspace, as defined by the singular value decomposition (SVD). Writing out summations explicitly, the SVD for a square matrix is:

$$(S)^{-1}_{ab} = \sum_{c'} U_{ac} s_c^{-1} U_{bc} \tag{23}$$

where the terms in the sum over c are discarded if s_c (the eigenvalues of $\mathbf{S}$, with eigenvectors $\mathbf{U}$) is zero. This inverse does not satisfy $\mathbf{S}^{-1}\mathbf{S}=\mathbf{1}$, but does yield the correct solution, $\mathbf{x}=\mathbf{S}^{-1}\mathbf{b}$ to linear equations $\mathbf{Sx}=\mathbf{b}$, if $\mathbf{b}$ lies within the range of $\mathbf{S}$. The SVD is also general for rectangular matrices, as needed for transformations between redundant and nonredundant representations.

Generalized inverses of the metrics for projected atomic orbitals are the contravariant metric matrices discussed above: the density matrix and its orthogonal complement respectively:

$$g^{ij} = \left\langle \phi^i \mid \phi^j \right\rangle = \delta^i_\mu P^{\mu\nu} \delta^j_\nu \tag{24}$$

$$g^{ab} = \left\langle \phi^a \mid \phi^b \right\rangle = \delta^a_\mu \left(g^{\mu\nu} - P^{\mu\nu} \right) \delta^b_\nu \tag{25}$$

This is proved either by considering the overlap of the contravariant projected functions themselves or by the fact that the product of these two matrices operating on any vector lying entirely within the occupied space yields the same function unmodified.

The contraction of the product of covariant and contravariant metrics yields an idempotent matrix which represents the Kronecker delta. For example, in the occupied case, we obtain

$$g^{ik} g_{kj} = \left\langle \phi^i \mid \phi_j \right\rangle = g^i_{\bullet j} \neq \delta^i_j \tag{26}$$

Thus redundant contravariant functions are not biorthogonal with redundant covariant functions. In the natural representation, P and Q must then be written as:

$$\hat{P} = |\phi_i\rangle \, g^i_{\bullet j} \, \langle \phi^j | \tag{27}$$

$$\hat{Q} = |\phi_a\rangle \, g^a_{\bullet b} \, \langle \phi^b | \tag{28}$$

We next consider rectangular coordinate transformations, $C^{j'}_{\bullet i}$, from a set of projected atomic orbitals, (19), (primed indices) to a linearly independent orthogonal set (unprimed indices). This transformation matrix is a set of molecular orbital coefficients. The generalized completeness relation above defines **C** and its generalized inverse **D** directly:

$$\big|\phi_i\big\rangle = \big|\phi_{j'}\big\rangle g^{j'}_{\bullet k'} \big\langle \phi^{k'} \mid \phi_i \big\rangle \equiv \big|\phi_{j'}\big\rangle C^{j'}_{\bullet i} \tag{29}$$

$$\big|\phi_{i'}\big\rangle = \big|\phi_j\big\rangle \big\langle \phi^j \mid \phi_{i'} \big\rangle \equiv \big|\phi_j\big\rangle D^j_{\bullet i'} = \big|\phi_j\big\rangle \big(C^j_{\bullet i'}\big)^* \tag{30}$$

D is both a left inverse and a right inverse. In the latter case, where **D** is undoing the transformation into the unprimed coordinates, the product of **C** and **D** is not the identity matrix, but rather the projector onto the occupied subspace in the mixed representation. Thus **D** works properly on all functions in the occupied subspace. We can also view **D** as a transformation from an orthogonal basis into the linearly dependent representation, with **C** as its generalized inverse.

This completes the demonstration that all basic operations of tensor algebra can be performed in terms of linearly dependent representations. Exact linear dependence evidently poses no practical or formal problem, because the space spanned by the linearly dependent functions remains well defined. Numerical complications can still arise from *near* linear dependence (where, for example, the eigenvalues of Equation (23) are very small, but not zero. In that case a threshold must be defined (the square root of machine precision for example) below which the eigenvalues and associated eigenvectors are discarded. Finally, while we have been discussing single-particle problems, it should also be emphasized that all of the results to be discussed in the next section on many-electron problems will also hold in the case of linear dependence, provided that the simple generalizations discussed here are applied.

3 Many-Electron Theory

We shall consider only so-called *single reference* theories of electron correlation. They begin from a single determinant of occupied orbitals, such as the mean-field Hartree-Fock solution, and treat electron correlation by substitutions of sets of occupied orbitals with virtual orbitals. As such, single reference methods are based upon an initial partitioning of the one-particle Hilbert space into two subspaces.

One consists of levels that are occupied in the reference determinant, and the other consists of unoccupied (virtual) functions:

$$\hat{I} = \hat{P} + \hat{Q} \tag{31}$$

The operators $\hat{P}$ and $\hat{Q}$ are projectors (as defined by idempotency, $\hat{P}^2 = \hat{P}$, $\hat{Q}^2 = \hat{Q}$) onto the occupied and virtual subspaces respectively. From Eq. (31) all functions which lie entirely within the occupied space are orthogonal to functions lying entirely in the virtual space (a one or two line proof). This applies to any mix of covariant and contravariant basis vectors, and is a condition of *strong orthogonality*. We shall adopt the convention that indices *i,j,k...* denote functions spanning the occupied subspace, indices *a,b,c...* denote functions spanning the virtual subspace, and indices μ,ν,λ,σ... denote functions lying in the full one-particle space. The presentation in this section generally follows our full paper on this problem[7].

3.1 Many-electron Determinants

We assume a single particle basis of dimension *N*, partitioned as in Eq. (31) into occupied and virtual subspaces (dimension n and N-n), each spanned by linearly independent functions (as discussed in the following section, linear dependence poses no practical problem either). The simplest many-electron wavefunction is a single determinant $|\Phi\rangle$, comprised of the n covariant functions spanning P. We call $|\Phi\rangle$ the *covariant reference ket*:

$$|\Phi\rangle = (n!)^{-1/2}\, det\left(|\phi_1\rangle\, |\phi_2\rangle \cdots |\phi_i\rangle \cdots |\phi_n\rangle\right) \tag{32}$$

What is the tensor character of $|\Phi\rangle$? This follows from how $|\Phi\rangle$ transforms under nonorthogonal transformations of the underlying single-particle basis functions spanning either P or Q. Since $|\Phi\rangle$ depends on tensor products of n single particle functions, it is an n-th rank antisymmetric tensor. The properties of antisymmetric tensors are well known as part of exterior algebra or Grassman algebra. An important fact is that if the dimension of the underlying single particle basis is S, then the dimension of the subspace of nth rank tensors which are antisymmetric is ${}^{S}\mathrm{C}_{n}$. This is 1 if S=n, as is the case for $|\Phi\rangle$ under transformations of the occupied space, or ${}^{N-n}\mathrm{C}_{0}$ for the virtual space. This permits substantial simplification, because a basis function spanning a one-dimensional subspace behaves much like a scalar under transformation of the basis.

In fact, a nonsingular linear transformation, $\mathbf{C}$, of the one-electron kets spanning the occupied space only alters the covariant reference ket by a constant, det($\mathbf{C}$). The proof is based on rewriting Eq. (32) as $|\Phi\rangle = (n!)^{-1/2}\, det\,(\Theta)$, where Θ is a matrix whose rows are electron labels, and whose columns are occupied covariant orbital labels. Under transformation with $\mathbf{C}$:

$$(n!)^{1/2}\, |\Phi\rangle' = det\,(\Theta') = det\,(\Theta C) = det\,(\Theta)\, det\,(C) = |\Phi\rangle\, det\,(C) \tag{33}$$

An antisymmetric n-th rank tensor which transforms this way is a *relative tensor* of rank 0 and weight 1 (or a pseudoscalar). Hence we can focus only on (the

small number of) single particle levels that are altered from their configuration (occupied or not) in the reference by the correlation treatment, rather than treating substituted determinants as general n-th rank tensors. The weight 1 refers to occupied basis transformations, while the tensor is of weight 0 with respect to virtual basis transformations, as is trivially proven.

Many electron expansion spaces are generated by replacing covariant occupied levels by covariant virtual levels. The set of all single substitutions of occupied levels, i, by virtual levels a, the corresponding set of all double substitutions, etc comprise well-defined subspaces of the full n-electron Hilbert space, $\mathbb{Q}_n$ (which consists of all substitutions through n-fold).

$$\mathbb{Q}_n = 0 \oplus S \oplus D \oplus T \oplus \cdots \oplus n \tag{34}$$

In general, electron correlation techniques seek an approximate solution of the Schrödinger equation that is defined only within such subspaces. For example, a trial wavefunction that depends on single and double substitutions can be completely determined by solving a Schrödinger-like equation within the $0 \oplus S \oplus D$ subspace.

We state the tensor properties of the substituted determinants that span these many-electron S, D, T subspaces, under transformations of the occupied and virtual single particle basis sets without proof. A covariant virtual level a, introduced by substitution, transforms covariantly. The hole index, i, of an occupied level that has been substituted, transforms contravariantly, consistent with creating a hole by projection with a biorthogonal contravariant bra orbital. These n-th rank antisymmetric tensors are also relative tensors of rank (m,m), for m-fold substitutions, with weight 1 for occupied basis transformations, and weight 0 with for virtual basis transformations. A single substitution of an occupied (covariant) level i by a virtual (covariant) level a is written as $\big|\Phi^{i}_{\bullet a}\big\rangle$, while double substitutions are $\big|\Phi^{ij}_{\bullet\bullet ab}\big\rangle$.

The space spanned by any given level of substitutions of occupied levels by virtuals is closed under separate nonsingular linear transformations of occupied and/or virtual levels amongst themselves. This follows directly from the fact that strong orthogonality between occupied and virtual subspaces is preserved under nonunitary (nonsingular) transformations of the one-particle basis set within the occupied and virtual subspaces. To give a concrete example, single substitutions are not mixed with double substitutions under transformations of this type. Note that subsets of these spaces are generally not closed under linear transformations.

The tensor character of substitution amplitudes is opposite to the many-electron basis vectors so that their contractions are properly invariant to transformations of the basis. Operators which generate invariant mixtures of substituted determinants can then be written as

$$\hat{T}_1\big|\Phi\big\rangle = t^{a}_{\bullet i}\big|\Phi^{i}_{\bullet a}\big\rangle \tag{35}$$

$$\hat{T}_2\big|\Phi\big\rangle = \tfrac{1}{4} t^{ab}_{\bullet\bullet ij}\big|\Phi^{ij}_{\bullet\bullet ab}\big\rangle \tag{36}$$

The amplitude tensors can be transformed from this natural representation with metric matrices in order to alter the character of the indices if desired.

3.2 Many-electron Bras and Slater's Rules

We define many-electron dual spaces, which preserve the biorthogonality present between contravariant and covariant single particle levels, starting from a determinant of contravariant bras, which is the n-electron *contravariant reference bra*:

$$\langle\Phi| = (n!)^{-1/2} det(\langle\phi^1|\langle\phi^2|\cdots\langle\phi^i|\cdots\langle\phi^n|) \tag{37}$$

There are corresponding manifolds of substituted bras, $\{\langle\Phi^a_{\bullet i}|\}$, $\{\langle\Phi^{ab}_{\bullet\bullet ij}|\}$ etc. *In the dual space, the tensor character of indices is reversed: occupied indices (holes) are covariant, virtual indices (particles) are contravariant.* For example, we replace occupied level $\langle\phi^i|$ with virtual level $\langle\phi^a|$ in the single substitution $\langle\Phi^a_{\bullet i}|$. These are all relative tensors of weight -1, because by construction they transform in a reciprocal way to the relative tensors of Sec. 3.1. They are the adjoints of the corresponding covariant many-electron kets.

Solution of Schrödinger-like equations in the many-electron subspace spanned by a given level of substituted determinants can be achieved by forming projection equations with the appropriate dual vectors. Biorthogonal representations arise in unsymmetric eigenvalue problems, such as coupled cluster excited state methods[19]. They are also sometimes employed in valence bond methods[20]. In our case, the projection equations are components of an absolute tensor which represents the many-electron equations in a subspace. It is invariant to transformations of the underlying basis, provided the projections are complete within each given substituted manifold (i.e. all double substitutions, for example).

Matrix elements of operators between many-electron bras and kets in the natural representation are slight generalizations of Slater's rules for the orthogonal case, with all new information contained within the covariant and contravariant character of the indices. The key to this close analogy is the natural representation in which covariant and contravariant character is consistently reversed between many-electron kets and bras. This is a tremendous contrast relative to the complexity of matrix elements between determinants of nonorthogonal functions in both bra and ket, which do not have the biorthogonality property.

For the identity operator (overlaps), as a result of orthonormality of single particle functions in the dual space with functions in normal (covariant) basis, only diagonal overlaps are nonzero. Thus:

$$\langle\Phi\ |\Phi\rangle = 1 \tag{38}$$

$$\langle\Phi^b_{\bullet j}\ |\ \Phi^i_{\bullet a}\rangle = \delta^i_j\delta^b_a \tag{39}$$

$$\langle\Phi^{cd}_{\bullet\bullet kl}\ |\ \Phi^{ij}_{\bullet\bullet ab}\rangle = (\delta^i_k\delta^j_l - \delta^i_l\delta^j_k)(\delta^c_a\delta^d_b - \delta^d_a\delta^c_b) = \delta^{ij}_{kl}\delta^{cd}_{ab} \tag{40}$$

In Eq. (40), we have simplified the final expression by introducing the generalized Kronecker delta (whose definition is obvious from the full equation). Nevertheless it is apparent from Eq. (40) that doubly substituted contravariant bras are not biorthogonal to doubly substituted covariant kets as we have defined them. This situtation is exactly the same as applies in the familiar orthonormal representation, and is a result of using a set of double substitutions which is redundant by roughly a factor of 4 (the permutationally related substitutions $\left|\Phi^{ij}_{\bullet\bullet ab}\right\rangle$, $\left|\Phi^{ji}_{\bullet\bullet ab}\right\rangle$, $\left|\Phi^{ij}_{\bullet\bullet ba}\right\rangle$, and $\left|\Phi^{ji}_{\bullet\bullet ba}\right\rangle$ are all equivalent). The somewhat cumbersome process of spin-adaptation can be used to define sets of orthonormal double substitutions, but we shall not pursue that issue here, preferring to use general spin orbitals. For a clear introductory discussion of how to proceed for the closed shell singlet case, see the article by Knowles, Schütz and Werner elsewhere in this volume.

For spin-orbital matrix elements involving one-electron operators, we derive results whose form is also familiar, as a direct result of the biorthogonality between many-electron basis kets and the dual basis bras:

$$\langle \Phi \,|\, F \,|\, \Phi \rangle = F^{i}_{\bullet i} = g^{ij} F_{ji} \tag{41}$$

$$\langle \Phi^{a}_{\bullet i} \,|\, F \,|\, \Phi \rangle = F^{a}_{\bullet i} = g^{ab} F_{bi} \tag{42}$$

We next consider two-electron operators. The tensor analog of the antisymmetrized ("double bar") two-electron integrals is a fourth rank tensor that in the natural representation has two covariant (ket) indices and two contravariant (bra) indices. There is also a corresponding fourth rank tensor for the parent ("single bar") two electron integrals. The definition of the antisymmetrized two electron integrals is:

$$\text{I\!I}^{pq}_{\bullet\bullet rs} = \langle \Phi^p \Phi^q \,|\Phi_r \Phi_s\rangle - \langle \Phi^p \Phi^q \,|\Phi_s \Phi_r\rangle \equiv I^{pq}_{\bullet\bullet rs} - I^{pq}_{\bullet\bullet sr} \tag{43}$$

This appears in matrix elements of the two-electron repulsion operator:

$$\langle \Phi \,|\, G \,|\, \Phi \rangle = \tfrac{1}{2} \text{I\!I}^{ij}_{\bullet\bullet ij} = \tfrac{1}{2} g^{ik} g^{jl} \text{I\!I}_{klij} \tag{44}$$

$$\langle \Phi^{a}_{\bullet i} \,|\, G \,|\, \Phi \rangle = \text{I\!I}^{aj}_{\bullet\bullet ij} = g^{ab} g^{ik} \text{I\!I}_{bkij} \tag{45}$$

$$\left\langle \Phi^{ab}_{\bullet\bullet ij} \middle| G \middle| \Phi \right\rangle = \text{I\!I}^{ab}_{\bullet\bullet ij} = g^{ac} g^{bd} \text{I\!I}_{cdij} \tag{46}$$

In addition to the natural representation, we have employed the metric tensors to re-express the results in terms of fully covariant two-electron integrals.

3.3 Second Quantization

In first quantization in the previous section, we have continued to develop a strong isomorphism between the algebra of quantum mechanics in orthogonal single particle basis sets, and nonorthogonal single particle basis sets when the natural representation is employed. It is therefore at least intuitive that other powerful tools of many-body electronic structure theory with orthonormal single particle basis sets will also carry over virtually unchanged to the nonorthogonal case provided the natural representation is employed. In particular, the tools of second quantization, and the associated diagrammatic methods for systematically treating the relatively complex algebra of many body theories is what we will want to use in general.

The purpose of this section is to describe the manner in which second quantization carries over from the case of an orthonormal single particle basis to the nonorthogonal case in the natural representation. We shall concentrate on the basic concepts, carrying them as far as necessary to show that Wick's theorem, the underpinning of diagrammatic methods, goes over unaltered. It is beyond our present scope to provide any real introduction to the use of Wick's theorem in evaluating many-body matrix elements. Rather, the key is to establish the manner in which standard textbook treatments (see for example refs. 2,3) of second quantization and diagrammatic methods can be directly employed in problems involving nonorthogonal single particle basis sets.

Due to our focus on the many-particle problem, it is not convenient to define creation and destruction operators with respect to the zero particle vacuum, unlike some previous work on nonorthogonal basis sets. Therefore, *we take the so-called quasiparticle (QP) vacuum as our reference*: a single determinant of n occupied functions. In the natural representation, the covariant reference ket contains the occupied set of covariant basis vectors, while the contravariant reference bra contains the set of occupied contravariant functions. *Thus for substituted many-electron kets, particles (levels outside the QP reference which become filled) are covariant in character while holes (levels within the QP reference which become vacated) are contravariant. In the dual space of substituted many-electron bras, particles are contravariant while holes are covariant.* All QP creation and destruction operators defined below obey this simple convention.

We must distinguish operations on virtual (unoccupied) functions (which can create or destroy particles in the QP vacuum) from operations on occupied functions (which will create or destroy holes in the QP vacuum). For the particle operators in the virtual space in the natural representation:

$$a_a^\dagger \equiv \text{QP particle creation operator for } |\phi_a\rangle \equiv \text{QP destruction operator for } \langle\phi^a|$$

$$a^b \equiv \text{QP particle destruction operator for } |\phi_b\rangle \equiv \text{QP creation operator for } \left\langle\phi^b\right|$$

The anticommutation relations for the particle operators follow by any of the usual textbook derivations as:

$$\left\{a^a, a_b^\dagger\right\} = \delta_b^a \tag{47}$$

$$\left\{a_a^\dagger, a_b^\dagger\right\} = \left\{a^a, a^b\right\} = 0 \tag{48}$$

As we have already seen in several other contexts, all differences relative to the usual orthogonal basis treatments are contained in the covariant and contravariant character of the indices. Of course this is only the case in the natural representation. Otherwise we would find overlaps for the anticommutator in Eq. (47). In a redundant representation, such as was discussed in the section on Linear Dependence, the idempotent matrix of overlaps of covariant and contravariant levels replaces the Kronecker delta:

$$\left\{a^a, a_b^\dagger\right\} = g^a_{\bullet b} \tag{49}$$

For functions spanning the occupied space, the identity of creation and destruction operators reverses, as we create and destroy holes in the QP vacuum:

$a^i \equiv$ QP hole creation operator for $|\phi_i\rangle \equiv$ QP hole destruction operator for $\left\langle \phi^i \right|$

$a_i^\dagger \equiv$ QP hole destruction operator for $|\phi_i\rangle \equiv$ QP hole creation operator for $\left\langle \phi^i \right|$

The anticommutation relations are:

$$\left\{a^i, a_j^\dagger\right\} = \delta^i_j \tag{50}$$

$$\left\{a_i^\dagger, a_j^\dagger\right\} = \left\{a^i, a^j\right\} = 0 \tag{51}$$

Due to strong orthogonality between levels in the occupied and virtual one-particle spaces, all anticommutators between particle and hole operators are zero, as in the orthogonal basis. In a redundant representation, $g^i_{\bullet j}$ replaces δ^i_j , analogous to how we showed above that Eq. (49) generalizes Eq. (47).

Contraction is central to evaluating matrix elements within second quantization. The contraction of operators x_1 and x_2 is defined as the difference between their product, and their QP normal ordered product. The *normal ordered* string is arranged such that all particle-hole creation operators are to the left, and all destruction operators are to the right, and is multiplied by the *parity* of the permutation (-1 for an odd number of pairwise swaps; $+1$ for an even number) necessary to take the product to normal order. Thus the contraction of a pair of operators (denoted by the overbar) is:

$$\overline{x_1 x_2} = x_1 x_2 - N\left[x_1 x_2\right] = \langle \Phi |\, x_1 x_2 \,| \Phi \rangle \tag{52}$$

where the N signifies that the following operator string is in normal order. Based on the anticommutation relations above, the only nonzero contractions between any pair of particle-hole creation and destruction operators are:

$$\overbrace{a^a a_b^\dagger} = \delta_b^a \tag{53}$$

$$\overbrace{a_i^\dagger a^j} = \delta_i^j \tag{54}$$

To evaluate many-electron operator matrix elements, we apply Wick's theorem (e.g. refs. 2,3 for orthogonal basis sets), which expresses a string of creation and destruction operators of length m, as a terminating sum of partial contractions of the string in normal order:

$$x_1 x_2 \cdots x_m = N\left[x_1 x_2 \cdots x_m\right] + \sum_{\substack{\text{single}\\ \text{contractions}}} N\left[x_1 \overbrace{x_2 \cdots} x_m\right] + \sum_{\substack{\text{double}\\ \text{contractions}}} N\left[x_1 \overbrace{x_2 \cdots} \overbrace{\cdots} x_m\right] + \ldots + \sum_{\substack{\text{m/2 fold}\\ \text{contractions}}} N\left[\overbrace{x_1 x_2}\ \overbrace{x_3 x_4}\ \cdots \overbrace{x_m}\right] \tag{55}$$

Since vacuum expectation values of normal-ordered operator strings are zero, the vacuum expectation value involves only the fully contracted terms. Again, apart from the newly explicit tensor character of the operators, these results are identical with the familiar orthogonal basis case. Relative to previous treatments of many-electron theory involving nonorthogonal single particle expansion spaces, this is the real value of the present approach.

Using either Wick's theorem directly, or from inspection of the orthogonal basis results, we can now obtain general expressions for one and two-electron operators:

$$\hat{H} = H_{\bullet i}^{i} + H_{\bullet q}^{p} N\left[a_p^\dagger a^q\right] \tag{56}$$

$$\hat{G} = \tfrac{1}{2} \mathit{\Pi}_{\bullet\bullet ij}^{ij} + \mathit{\Pi}_{\bullet\bullet qi}^{pi} N\left[a_p^\dagger a^q\right] + \tfrac{1}{4} \mathit{\Pi}_{\bullet\bullet rs}^{pq} N\left[a_p^\dagger a_q^\dagger a^s a^r\right] \tag{57}$$

In these expressions, indices *i,j,k..* are occupied levels, and *p,q,r,s* are general levels where the indices run over both occupied and virtual ranges, and the antisymmetrized two-electron integrals were defined in Eq. (43). These second quantization tools are sufficient to reduce n-electron matrix elements to mixtures of 1 and 2-electron matrix elements, as for the generalized Slater's rules given in the previous section, or the correlation methods discussed in the sections below.

4 Nonorthogonal Functions for Local Electron Correlation

The development of what may be called "fast methods" for evaluating electron correlation is a problem of both fundamental and practical importance, because of these unphysical increases in computational complexity with molecular size which afflict "exact" implementations of electron correlation methods. The purpose of

this part of the chapter is to discuss our recent progress towards developing a new family of local correlation models that are based on atomic truncations[21,22,23]. Our emphasis here is on the formulation of the theory, as an interesting and topical application of the general tensor approach to the many-body problem developed in the first part of this chapter. We restrict our discussion entirely to second order Møller-Plesset (MP2) theory[24], the simplest useful wavefunction-based electron correlation method. Local models that perform effectively at the MP2 level are expected to be transferable to other theories of electron correlation in terms of single and double substitutions, such as coupled cluster theory with single and double substitutions (CCSD). Our local MP2 models are also an interesting starting point for treating triple substitutions (as needed for example in MP4 and CCSD(T)) as we briefly describe in Section 5.1.

The MP2 method was originally proposed in 1934[24]. Revived in the mid-1970's, it remains highly popular today, because it offers systematic improvement in optimized geometries and other molecular properties relative to Hartree-Fock theory[25]. Indeed, in a recent comparative study of small closed shell molecules[26] , MP2 outperformed the much more expensive CCSD method for such properties! Relative to state-of-the-art Kohn-Sham density functional theory (DFT) methods[27] , which are the most economical methods to account for electron correlation effects, MP2 has the advantage of properly incorporating long-range dispersion forces. While its computational cost scales as the 5^{th} power of molecular size, quite efficient semidirect methods[28] permit routine applications to medium-sized molecules.

We do not attempt to comprehensively review other efforts at developing fast methods for electron correlation. However, it is important to note that, to date, the most successful approach of this type is the method pioneered by Saebø and Pulay (SP)[10,29,30]. The SP model has now been adopted and extended by several groups, both at the MP2 level[31,32,33,34] and also at the level of self-consistent treatment of single and double substitutions[35,36,37]. We shall later compare the SP method against our new methods. Two important differences are that the SP method does not yield continuous potential energy surfaces, and sometimes fails to treat symmetry equivalent atoms on an equal footing, as in benzene for example. However, to this stage the SP method has advanced much further as far as efficient implementation is concerned. An excellent discussion of the current state of the art is given in the chapter by Knowles, Schütz and Werner in this volume.

As a preliminary, we discuss how local correlation impacts theoretical and numerical modeling in quantum chemistry. Generally, it is useful to distinguish two main classes of errors in an electronic structure calculation:

(1) *Intrinsic model errors*, due to employing incomplete one-particle basis sets, and incomplete descriptions of the many particle expansion spaces associated with the one particle basis. Such errors may be large, but by virtue of understanding the chemistry of the model and systematically studying its behavior over a range of applications, one can understand the different conditions under which it may produce reliable versus unreliable results. The theoretical model should be constructed in as clean a fashion as possible, so that successes and failures of the model can be cleanly understood.
(2) *Numerical error*, due to not exactly implementing the model. There is a strong

argument for attempting to keep this source of error as small as possible. Without doing so, one cannot assess the intrinsic model errors cleanly. Furthermore, the reproducibility of calculations is impeded if there is significant numerical noise. However, in the context of fast methods, the use of looser thresholds and cutoffs leads to significant speedups. Therefore it is advisable to control precision carefully because it is strongly correlated with computational cost.

Ideally, the development of fast methods for treating electron correlation should not impact either model errors or numerical errors. Unfortunately this is not possible at present, as may be appreciated from the following rough argument. *Spatial locality* is what permits reformulations of electronic structure methods that yield the same answer as traditional methods, but faster. The one-particle density matrix decays exponentially with a rate that relates to the HOMO-LUMO gap in periodic systems. From model studies, as well as formal considerations[17] , the decay rates are fairly well understood. When length scales longer than this characteristic decay length are examined, sparsity will emerge in both the one-particle density matrix and also pair correlation amplitudes expressed in terms of localized functions. Very roughly, such a length scale[9] is about 5 to 10 atoms in a line, for a good insulator such as alkanes. Hence sparsity emerges beyond this number of atoms in 1-d, beyond this number of atoms squared in 2-d, and this number of atoms cubed in 3-d. Thus for three-dimensional systems, locality only begins to emerge for systems of between hundreds and thousands of atoms.

If we wish to accelerate calculations on systems below this size regime, we must therefore introduce additional errors into the calculation, either as numerical noise through looser tolerances, or by modifying the theoretical model, or perhaps both. Our approach to local electron correlation is based on modifying the theoretical models describing correlation with an additional well-defined local approximation. We do not attempt to accelerate the calculations by introducing more numerical error because of the difficulties of controlling the error as a function of molecule size, and the difficulty of achieving reproducible significant results.

From this perspective, local correlation becomes an integral part of specifying the electron correlation treatment. This means that the considerations necessary for a correlation treatment to qualify as a well-defined theoretical model chemistry[25] apply equally to local correlation modeling. The approximations should be:

a) *Size-consistent*: meaning that the energy of a supersystem of two noninteracting molecules should be the sum of the energy obtained from individual calculations on each molecule.
b) *Uniquely defined:* Require no input beyond nuclei, electrons, and an atomic orbital basis set. In other words, the model should be uniquely specified without customization for each molecule.
c) *Yield continuous potential energy surfaces:* The model approximations should be smooth, and not yield energies that exhibit jumps as nuclear geometries are varied.

While these criteria are simple and indeed almost trite to state, they are quite easy to violate, and indeed many procedures for performing efficient calculations of electron correlation effects do not meet the 2^{nd} and 3^{rd} criteria. For example,

in the SP local correlation method, the choice of which substitutions to retain depends on the nuclear geometry[38,39]. Therefore when consistently applied the resulting potential energy surfaces are not smooth, unless the substitutions to keep are defined at a single arbitrary geometry. Similarly the configuration selection involved in current implementations of Nakatsuji's SAC-CI method yields results that depend on geometry, and this either causes potential energy surfaces to be non-smooth or the results must depend on the reference geometry chosen[40]. In the SP method, other difficulties arise in cases where the molecular orbitals do not localize cleanly[41], leading to difficulties in the consistent treatment of molecules like benzene, and to problems in defining the substitutions to retain when treating transition structures. In the new models described later in this part of this section (Section 4.3 and onwards), the requirements of a theoretical model chemistry will be strictly satisfied, by discarding many degrees of freedom according to an atomic criterion.

4.1 Second Order Mller-Plesset Theory

First, let us begin by discussing the equations of our chosen correlation method, MP2 theory, in general terms. A key result from earlier in this chapter is that use of the natural representation leads to equations for electron correlation methods that are isomorphic to traditional orthogonal basis derivations, but substantially more general. This is what we shall illustrate in this section for MP2 theory. The usual formulation of MP2 theory involves the assumption of the canonical molecular orbitals (so that the zero order Hamiltonian is diagonal in the many-electron basis). However, more general "noncanonical" forms of MP2 theory which are invariant to unitary transformations within the occupied and virtual spaces were introduced in the context of gradient theory, and open shell problems[42]. With Pulay and Saebø's local MP2 method[10], a yet more general form was presented, in which nonorthogonal functions were employed to span the virtual function space. Our tensor formalism[7] yields the most general MP2 equations possible, as they are invariant to nonunitary transformations within the occupied and virtual spaces individually.

The basis of Møller-Plesset perturbation theory is partitioning the full Hamiltonian into a zero order (mean field) part, and a first order component which describes the fluctuations associated with electron correlation effects. This may be written in second quantized form as:

$$\hat{H} = \hat{F}^{(0)} + \lambda \hat{V}^{(1)} \tag{58}$$

$$\hat{F}^{(0)} = F^{i}_{\bullet i} + F^{a}_{\bullet b} N\left[a^{\dagger}_{a} a^{b}\right] - F^{i}_{\bullet j} N\left[a^{j} a^{\dagger}_{i}\right] \tag{59}$$

$$\hat{V}^{(1)} = -\tfrac{1}{2} \mathbb{I}^{ij}_{\bullet\bullet ij} + F^{i}_{\bullet a} N\left[a^{\dagger}_{i} a^{a}\right] + F^{a}_{\bullet i} N\left[a^{\dagger}_{a} a^{i}\right] + \tfrac{1}{4} \mathbb{I}^{pq}_{\bullet\bullet rs} N\left[a^{\dagger}_{p} a^{\dagger}_{q} a^{s} a^{r}\right] \tag{60}$$

F is the Fock operator corresponding to the reference single determinant wavefunction, and, as usual indices *i,j,k...* are occupied, and *a,b,c...* are virtual. The

anti-symmetrized two electron integrals were previously defined in Eq. (43). By placing occupied-virtual elements of the Fock operator into the fluctuation potential (Equation (60)), the single reference is an eigenket of Equation (59) even when these elements are nonzero. The matrix elements and creation and destruction operators entering Equations (59) and (60) have tensorial significance.

Performing perturbation expansions of the full wavefunction together with the Hamiltonian leads to conditions for the first order wavefunction. These first order equations for the perturbed wavefunction may be resolved by projection with singly and doubly substituted bra determinants (in the dual basis of course) to yield equations for the first order single and double substitution amplitudes (contained within the corresponding substitution operators, $\hat{T}_1^{(1)}$and $\hat{T}_2^{(1)}$):

$$\left\langle \Phi^a_{\bullet i} \left| \hat{F}^{(0)} \left| \hat{T}_1^{(1)} \Phi \right. \right\rangle + \left\langle \Phi^a_{\bullet i} \left| \hat{V}^{(1)} \right| \Phi \right\rangle = 0 \tag{61}$$

$$\left\langle \Phi^{ab}_{\bullet\bullet ij} \left| \hat{F}^{(0)} \left| \hat{T}_2^{(1)} \Phi \right. \right\rangle + \left\langle \Phi^{ab}_{\bullet\bullet ij} \left| \hat{V}^{(1)} \right| \Phi \right\rangle = 0 \tag{62}$$

The substituted bras represent the replacement of contravariant occupied functions by contravariant virtual functions.

These equations may be written out in explicit spin-orbital notation quite easily, exploiting the direct analogy between the natural representation and orthogonal basis techniques.

$$F^a_{\bullet b} t^b_{\bullet i} - F^j_{\bullet i} t^a_{\bullet j} + F^a_{\bullet i} = 0 \tag{63}$$

$$F^b_{\bullet c} t^{ac}_{\bullet\bullet ij} + F^a_{\bullet c} t^{cb}_{\bullet\bullet ij} - F^k_{\bullet j} t^{ab}_{\bullet\bullet ik} - F^k_{\bullet i} t^{ab}_{\bullet\bullet kj} + \mathbb{I}^{ab}_{\bullet\bullet ij} \tag{64}$$

Here $\boldsymbol{t}$ are the unknown first order amplitudes, that are now defined in terms of two electron integrals, and matrix elements of the Fock operator.

Finally, the second order expansion of the energy is obtained from the second order expansion of the eigenvalue equation followed by projection with the bra determinant:

$$E^{(2)} = \left\langle \Phi \left| \hat{V}^{(1)} \right| \left(\hat{T}_1^{(1)} + \hat{T}_2^{(1)} \right) \Phi \right\rangle = F^i_{\bullet a} t^a_{\bullet i} + \tfrac{1}{4} \mathbb{I}^{ij}_{\bullet\bullet ab} t^{ab}_{\bullet\bullet ij} \tag{65}$$

Equations (63), (64), and (65) embody all the invariances inherent in MP2 theory: they are invariant to general nonunitary linear transformations within the occupied and virtual subspaces respectively. These equations, despite their perturbational nature, must be solved iteratively because of the off-diagonal Fock terms in Equations (63) and (64). They can be specialized to the textbook[1] canonical case by:

(1) assuming an orthogonal basis so that the covariant-contravariant distinction vanishes),
(2) satisfying the Brillouin theorem (such that $F^a_{\bullet i} = 0$) so that the first order singles are zero, and,

(3) diagonalizing the occupied-occupied and virtual-virtual blocks of the Fock matrix (so that the doubles equations uncouple).

Via the use of metric matrices, it is straightforward to alter the tensor character of the integrals and amplitudes that enter the above equations. As mentioned earlier, the natural representation is particularly convenient for defining the nonorthogonal equations of electron correlation methods, but it is not necessarily the form that we will subsequently use. Rather, since all one and two electron matrix elements are made directly in the given covariant basis, we prefer to use the *covariant integral representation*, which treats all amplitudes as contravariant.

Let us recast the MP2 equations given above into the covariant integral representation. Acting with metric matrices to convert all indices in the matrix elements to covariant and all indices in the amplitudes to contravariant leads to the following alternative form:

$$E^{(2)} = F_{ia}t^{ai} + \tfrac{1}{4}\mathit{I}\!\mathit{I}_{ijab}t^{abij} \tag{66}$$

$$\big(F_{ab}g_{ij} - g_{ab}F_{ij}\big)t^{bj} + F_{ai} = 0 \tag{67}$$

$$\big[\big(F_{ac}g_{ik} - g_{ac}F_{ik}\big)g_{bd}g_{jl} + g_{ac}g_{ik}\big(F_{bd}g_{jl} - g_{bd}F_{jl}\big)\big]t^{ckdl} + \mathit{I}\!\mathit{I}_{aibj} \tag{68}$$

Note that in writing Equations (66) and (68) we have chosen to reorder the indexes in the integral and amplitude tensors to group occupied and virtual indexes together. This is in preparation for making local approximations based on keeping only occupied and virtual functions that share a common atom: the atomic single substitution basis that is discussed in more detail later on.

4.2 The Nonorthogonal Orbitals

As discussed in the introduction to this section, we have decided to focus on atom-centered functions, so that our new models for local correlation can be based on *atom-centered selection* of the significant pair correlation amplitudes. The use of atom-based functions is very common within single electron theories of electronic structure such as Hartree-Fock theory and density functional theory. In the context of the tensor-based electron correlation formalism, we require sets of functions spanning the occupied and virtual subspaces.

The simplest approach is to use projected atomic orbitals: projected into the occupied space and the virtual space respectively:

$$|\phi_i\rangle = \hat{P}\,|\phi_\mu\rangle\,\delta_i^\mu = |\phi_\mu\rangle\,P_{\bullet\nu}^{\mu}\delta_i^\nu \tag{69}$$

$$\big|\phi_a\big\rangle = \hat{Q}\big|\phi_\mu\big\rangle\delta_a^\mu = \big(\hat{I} - \hat{P}\big)\big|\phi_\mu\big\rangle\delta_a^\mu = \big|\phi_\mu\big\rangle\big(\delta_a^\mu - P_{\bullet\nu}^{\mu}\delta_a^\nu\big) \tag{70}$$

These functions are assigned to atoms by the atom label of the parent atomic orbital, thus partitioning the redundant functions amongst the atoms according to

the number of AO's on each atom in a manner that is well-defined for any atomic arrangement. The resulting functions are exponentially localized for insulators, as they have the decay properties of the one-particle density matrix **P** itself. The use of functions of this type is satisfactory for the virtual space and was pioneered successfully in the earlier local correlation models of Pulay and Saebø. However projected atomic orbitals are highly redundant (and hence extremely inefficient) for the occupied space. Increasing the number of basis functions per atom would increase the number of functions spanning the occupied space, even though the dimension (rank) of the occupied space itself is unaltered!

The simplest atom-centered basis that is capable of spanning the occupied space is a *minimal basis* of core and valence atomic orbitals on each atom. Such a basis is necessarily redundant because it also contains sufficient flexibility to describe the empty valence antibonding orbitals necessary to correctly account for nondynamical electron correlation effects such as bond-breaking. Atom-optimized minimal basis sets (e.g. of the STO-nG type) are of course notoriously poor in practical calculations. Yet it is still quite possible to define a localized minimal basis that spans the occupied space at the end of a large basis set calculation. We term such functions *extracted polarized atomic orbitals* (EPAO's). It is even possible to explicitly perform an SCF calculation in terms of a molecule-optimized minimal basis of *polarized atomic orbitals* (PAO's). These two approaches each have strengths and weaknesses, and we discuss them in turn below.

First we summarize the situation. The number of functions spanning the occupied subspace will be the minimal basis set dimension, M, which is greater than the number of occupied orbitals, O, by a factor of up to about 2. The virtual space is spanned by the set of projected atomic orbitals whose number is the atomic orbital basis set size N, which is fractionally greater than the number of virtuals V=N-O. The number of double substitutions in such a redundant representation will be typically 3 to 5 times larger than the usual total. This will be more than compensated by reducing the number of retained substitutions by a factor of the number of atoms, A, in the local triatomics in molecules model, or a factor of A^2 in the diatomics in molecules model.

4.2.1 Extracted Polarized Atomic Orbitals

The *extracted polarized atomic orbitals* (EPAO's)[43] of a molecule are a minimal basis set that is derived subsequent to an extended basis electronic structure calculation. Thus the EPAO's are themselves linear combinations of AO's: $|\alpha\rangle = \sum_\mu B_{\alpha\mu} |\mu\rangle$. Based on numbers of core and valence atomic orbitals, the number of EPAO's for any given atom is 1 for H and He, 5 for first row elements, 9 for second row elements, etc. The set of EPAO's on each atom will be invariant to rotations amongst individual EPAO's on that atom. For simplicity, we will constrain the full set of EPAO's to form an orthonormal set, $\langle \alpha \mid \beta \rangle = \delta_{\alpha\beta}$.

EPAO's will be determined analogously to localized occupied orbitals in that an EPAO delocalization functional, L, is minimized with respect to variations of the atomic subspaces. Furthermore, these variations are subject to the constraint that the full space of EPAO's completely contains the one-particle density projector,

$\hat{\rho} = \sum_i |i\rangle \langle i|$, which can be abbreviated by the expression $\hat{\rho} \subset \hat{\tau}$, where the full space of EPAO's, $\hat{\tau}$, is the union of all the atomic subspaces in a system, $\hat{\tau} = \sum_A \hat{\tau}_A$.

Many definitions of the delocalization functional are possible, and a number have been explored. The choice we shall employ here is to minimize the second moment of the EPAO's, with the imposed condition that their centers coincide with the atomic centers:

$$L[\hat{\tau}_A, \hat{\tau}_B, ...] = \underset{\hat{\rho} \subset \hat{\tau}}{Min} \left\{ \sum_A k_A \sum_{\alpha,\beta \in A} \langle \alpha | \left(\hat{r} - z_A\right)^2 | \beta \rangle D_{\alpha\beta} \right\} \tag{71}$$

$D_{\alpha\beta} = \langle \alpha | \hat{\rho} | \beta \rangle$ is the one-particle density matrix evaluated in the EPAO basis. This occupancy weighting reduces the sensitivity of the results to the presence of empty or nearly empty orbitals in the minimum basis. The virtual part of the basis is still fully determined because the occupancy weighting is not a full projection into the occupied space. Equation (14) contains an atomic scaling factor, k_A, which we have included to make the definition of the localization criterion as general as possible.

A reasonable choice of the scaling factor k_A, will be to roughly normalize the radii of valence EPAO's of elements from different rows of the periodic table. We use a very simple scaling factor that achieves this objective, which is the inverse of the principal quantum number, n, of the atom. Thus, for elements H and He, the factor is 1; for elements Li through Ne, the factor is 1/2; for elements Na through Ar, the factor is 1/3, and so on. This choice very roughly mirrors trends in the sizes of the noble gas atoms, and has the merit of involving no adjustable parameters. Finally, the localized atom-centered occupied orbitals are obtained by projection of the EPAO's into the occupied space as $|i_A\rangle = \hat{P} |\alpha_A\rangle$. The projected functions remain atom-centered in general, but are linearly dependent as their number M is greater than the dimension of the occupied space, O.

4.2.2 Polarized Atomic Orbitals

Instead of extracting the minimal basis after the conclusion of an SCF calculation in an extended basis, can one directly perform the SCF calculation in terms of a molecule-optimized minimal basis set of polarized atomic orbitals? We have shown[44] that such an approach is feasible as long as a constraint is imposed on the form of the PAO's: namely that the PAO's are linear combinations only of the atomic orbitals on the corresponding atom. So, given a (large) secondary basis $\{|\phi_\mu\rangle\}$, the (small) basis of PAO's $\{|\phi_\alpha\rangle\}$ are defined via an atom-blocked (let PAO index α belong to atom A) rectangular transformation, $\mathbf{B}$:

$$|\phi_\alpha\rangle = \sum_{\mu \in A} |\phi_\mu\rangle \, B_{\mu\alpha} \tag{72}$$

The transformation coefficients are determined as part of the variational Hartree-Fock calculation, in which the density matrix or molecular orbitals are defined in terms of PAO's.

The PAO's appear very promising as a method for performing Hartree-Fock (HF) (or other self-consistent field methods such as density functional theory) calculations on large molecules. The computational requirements are reduced significantly because a PAO calculation does not require full matrices to be evaluated in the large atomic orbital basis. Additionally a PAO calculation by definition provides a minimal basis set of atom-centered functions that span the occupied space (after projection with the density operator).

It is important to emphasize that as a result of the atom-blocked constraint, the PAO-HF energy is necessarily slightly higher than the HF energy directly computed in the secondary atomic orbital basis. We have reported two studies of the magnitudes of these differences both for absolute and relative energies[44,45]. In terms of absolute energies, the error introduced by the PAO model is generally smaller than normal basis set incompleteness effects. Relative energies are quite well reproduced, such that the PAO model dramatically outperforms any conventional minimal basis set. Errors in absolute and relative energies can be reduced by employing larger secondary basis sets in the PAO calculation. Some preliminary studies of the convergence of results with size and composition of the secondary basis have been completed[45].

The appeal of PAO's for local MP2 calculations is that the atom-blocked form of the transformation to PAO's can potentially accelerate the first two steps of the atomic orbital to molecular orbital two-electron integral transformation. These steps are often rate-determining, particularly for large basis sets. We do not report any electron correlation calculations using PAO's in this chapter, but this subject is under active study in our laboratory[46].

4.3 Atoms in Molecules Models of Local Correlation

We are now at the stage where local models of electron correlation can be formulated and tested. The developments described in the previous sections constitute a formally exact re-expression of standard theories of electron correlation, using MP2 theory as the working example. In the traditional formulation, the nonredundant canonical molecular orbitals are employed to express many-electron theories. The canonical MO's are in general fully delocalized throughout the system. In our re-expression of electron correlation theories, redundant sets of atom-centered functions are used to express the working equations. This takes advantage of the fact that correlation theories are fundamentally independent of the choice of the functions used to span a given occupied and virtual space. As described above, the occupied space is spanned by a minimal basis set of distorted (polarized) atomic orbitals that are projected into the occupied space. This is the smallest set of atom-centered functions capable of performing this role. The virtual space is spanned by the full atomic orbital basis projected into the virtual space. As a result of using atom-centered functions, we can truncate the *overcomplete* set of double substitutions based on an atomic criterion, so as to satisfy the requirements of a theoretical model chemistry, as discussed at the start of Section 4.

At a conceptual level, we can distinguish four possible levels of truncation, based on increasingly severe atomic truncations:

a) *Tetra-atomics in molecules.* If no truncation (apart from numerical cutoffs) is employed, substitutions will generally occur from occupied atomic-like orbitals centered on two atoms to virtual atomic-like orbitals centered on two other atoms. As discussed above, based on sparsity considerations, we expect cross-overs relative to conventional MP2 will occur only for quite large systems. Again, this is the motivation for considering stronger, ansatz-based truncations.

b) *Triatomics in molecules* (TRIM) is the level of modeling where only double substitutions with one occupied and one virtual orbital on a common atom are retained. This immediately reduces the number of retained double substitutions to no more than A^3 in the number of atoms, A. Substitutions that are doubly ionic (and often identified with basis set superposition error) are eliminated, while singly ionic and covalent substitutions are retained.

c) *Diatomics in molecules* (DIM) is the case where one occupied-virtual substitution is restricted to be on a common atom, while the other occupied-virtual substitution is restricted to be on another atom. This eliminates all ionic substitutions, and so may alternatively be called the covalent ansatz[21,22]. The correct long-range dispersion force can still be recovered at this level of truncation.

d) *Atoms in molecules* is the case where the only retained double substitutions are those for which all occupieds and virtuals are on a common atom. Long-range dispersion is no longer recovered, and therefore we shall not consider this model further in the present context. However, it is of interest for treating nondynamical correlation associated with bond-breaking, as we discuss briefly in Section 5.2.

For practical purposes, there are then three levels of local correlation model treatment that may be distinguished in the context of MP2 theory: exact, triatomics, and diatomics. In the following sections, we shall develop the explicit expressions necessary to evaluate the energy associated with these models of local electron correlation.

4.4 Atomic Single Substitution Basis

Any set of orbital substitutions describing correlated fluctuations of electrons from a mean field single determinant starting point can be represented in terms of direct products of single substitutions. For example the set of double substitutions is the direct product of the space of single substitutions with itself, the set of triple substitutions is a 3-way direct product of the set of single substitutions, and so forth.

We have expressed the space of occupied functions in terms of a minimal basis of atom-centered functions, and the space of virtual (unoccupied) functions is spanned by a redundant set of projected atomic orbitals. This means that the set of single substitutions may now be divided into two categories:

(1) *Atomic Single Substitutions* (ASIS): These are single replacements where both the occupied function and the virtual function belong to the same atom.

(2) *Interatomic Single Substitutions*: Single replacements where the occupied function and the virtual function belong to different atoms.

The basis of our local truncations will be to retain only the ASIS basis for some (or all) of the single substitutions that enter into electron correlation theories. This approximation will express the locality inherent in electron correlation: namely excitations that involve transporting electrons between spatially different regions of a molecule are energetically insignificant. Retention of only the ASIS substitutions is actually not quite as drastic an approximation as it first appears. Since the functions spanning both the occupied and the virtual spaces are both nonorthogonal and linearly dependent, the ASIS space and the space spanned by the interatomic single substitutions are not disjoint. In fact the ASIS space includes a significant fraction of the excitations on neighboring (bonded) atoms.

The dimension of the ASIS set grows linearly with the number of atoms, with the coefficient being the average number of occupied-virtual product functions on an atom. By contrast, of course, the full set of single substutions grows quadratically with the number of atoms. Therefore we are going to reduce the number of amplitudes by one power of system size for each pair of occupied-virtual indices for which we make an ASIS approximation. Specifically, we can now symbolically write the form of both the DIM and TRIM models.

The full MP2 model involves the direct product of all single substitutions with themselves:

$$\{aibj\}_{full} = \{a\} \otimes \{i\} \otimes \{b\} \otimes \{j\} \tag{73}$$

The DIM model applies the atomic single substitution approximation to the two sets of single substitutions whose direct product defines the double substitutions. Denoting an ASIS substitution as (ai) where the use of the parenthesis reminds us that both occupied and virtual functions are centered on common atoms, we may write the space of DIM substitutions as:

$$\{aibj\}_{DIM} = \{(ai)\} \otimes \{(bj)\}\,. \tag{74}$$

Clearly electrons on pairs of atoms are correlated in this way, and, as discussed below, this is sufficient to correctly describe long-range dispersion interactions.

The TRIM model corresponds to applying the atomic single substitution approximation only to one of the two single substitutions whose direct product comprises the double substitutions of MP2 theory. To make such a model consistent the correlation space should be the union of the two possible ways of doing this. The TRIM set of double substitutions may then be written as follows:

$$\{aibj\}_{TRIM} = \{(ai)\} \otimes \{b\} \otimes \{j\} \oplus \{a\} \otimes \{i\} \otimes \{(bj)\}\,. \tag{75}$$

The TRIM model expands the DIM substitutions to explicitly include single electron transfers between atoms that can be separated. Full double substitutions, TRIM doubles and DIM doubles form a well-defined hierarchy of models for local correlation.

Roughly how do we expect the DIM and TRIM models to compare to the existing Saebø-Pulay (SP) model for local correlation? Or, alternatively, what does that model look like in terms of an expression like Equation (74) or (75)? If it were implemented using atomic truncation criteria, the set of retained double substitutions in the SP method could be expressed as follows:

$$\{aibj\}_{SP} = \{(ai)\} \otimes \{(bj)\} \oplus \{i\} \otimes \{(abj)\} \oplus \{(iab)\} \otimes \{j\}\,. \tag{76}$$

In Equation (76), quantities in parentheses are centered on a common atom. Thus the SP model can be expected to lie intermediate between our DIM and TRIM models. The first term in Equation (76) is the direct product of ASIS terms that are responsible for long-range correlation, and are included in the DIM model. The 2^{nd} and 3^{rd} terms are single electron transfers between pairs of atoms, in which charge is consolidated from one electron on the atom of orbitals i and j to two electrons on the atom of either orbital i or orbital j. These single electron transfers are a *subset* of those included in the TRIM model. TRIM additionally includes substitutions in which charge is separated from one atom to two atoms, as well as singly ionic transfers that couple three atoms together rather than two.

4.5 Orthogonalizing and Canonicalizing the ASIS Basis

Recall for a moment how the molecular orbitals are treated in conventional electronic structure theory. Normally they are first orthogonalized, which eliminates all overlap matrices from the equations, and then they are canonicalized, meaning that they are transformed to a basis which makes the Fock operator diagonal. As a result, the MP2 equations uncouple (for example, refer back to the steps listed after Equations (65), and the MP2 correlation energy can be obtained without iterations.

It is possible (and indeed, probably necessary) to do exactly the same things with the ASIS basis. We can orthogonalize, such that the overlap matrix (or metric) becomes diagonal. The ASIS overlap matrix is merely the direct product of the occupied and virtual overlaps. Denoting indexes in the ASIS basis as *K,L...* (they are equivalent to atomic (ai), (bj)... pairs), we can write:

$$G_{KL} \equiv G_{(ai)(bj)} = g_{ab} g_{ij} \tag{77}$$

This overlap matrix can be explicitly diagonalized (and thus inverted by singular value decomposition) quite readily for systems of up to 150 heavy atoms or so (the computational effort scales with the cube of molecule size). This yields the orthogonalizing transformation $\mathbf{G}^{-1/2}$. The orthogonalized ASIS basis will be called the OASIS basis.

Why the need for orthogonalizing the ASIS basis via singular value decomposition? The reason that explicit orthogonalization is probably necessary is that the condition number of the ASIS metric, Equation (77), is tremendously large, as a result of making the atomic truncation, and using nonorthogonal and linearly dependent functions. This means that while the matrix described by Equation (77) becomes sparse as the system size grows large, iterative methods to invert the metric will perform tremendously poorly. If we had not made the ASIS approximation

this would not necessarily be the case, since then we could invert separately the occupied and virtual metrics entering Equation (77).

The relevant energy matrix in the ASIS basis is evident in Equations (68), and is essentially a matrix of energy differences, which we shall denote as **D**:

$$D_{KL} \equiv D_{(ai)(bj)} = \left(F_{ac}g_{ik} - g_{ac}F_{ik}\right) \tag{78}$$

Canonicalizing this matrix requires two steps. First, **D** must be transformed into the OASIS basis (which we shall denote with overbars):

$$D_{\bar{K}\bar{L}} = \sum_{K,L} G^{-1/2}_{K\bar{K}} D_{KL} G^{-1/2}_{L\bar{L}} \tag{79}$$

The second step is to diagonalize **D** in this orthogonalized basis:

$$D_{\bar{K}\bar{L}} = \sum_{\bar{P}} U_{\bar{K}\bar{P}} E_{\bar{P}} U_{\bar{L}\bar{P}} \tag{80}$$

The product of these two transformations is the matrix **T** that transforms from the ASIS basis to the canonicalized OASIS (or COASIS) basis:

$$T_{K\bar{P}} = \sum_{K,\bar{K}} G^{-1/2}_{K\bar{K}} U_{\bar{K}\bar{P}} \tag{81}$$

4.6 Energy Expression

The MP2 spin-orbital expression for the energy that we shall use is:

$$E_{MP2} = \tfrac{1}{2} I_{aibj} t^{aibj} \tag{82}$$

where $I_{aibj} \equiv (ai\,|bj\,) = \langle ab\,|ij\,\rangle$ are unsymmetrized two-electron integrals. In Mulliken notation, such integrals correspond simply to the Coulomb interaction between the charge distribution corresponding to the product of virtual a and occupied i, and the charge distribution due to the product of virtual b and occupied j. The first order amplitudes are denoted as t^{aibj}. Relative to our expression in Equations (65) and (66), we are:

(a) Using unsymmetrized rather than symmetrized integrals. This leads to the factor of 1/2 instead of 1/4, and of course makes no difference in the absence of any local truncation. In the presence of local truncation, as discussed further below, Equation (82) is a preferable starting point.

(b) Omitting the singles substitution contribution to the energy. This is because, if present, we shall treat it exactly, and we wish to focus attention on the critical doubles contribution.

While Equation (82) is equivalent to the previous expressions in the absence of local truncations, it is important to emphasize that the result of making local truncations will be different depending on which starting point is used. This should

not be disturbing upon reflection: it is simply a consequence of the fact that local modeling breaks many of the invariances that are present in the original nonlocal equations. Equation (82) is a preferable starting point because it will yield 100% of the long-range correlation energy in the context of the DIM model. By contrast, using Equation (66) will yield only 50% of the long-range correlation energy under DIM truncation!

4.7 Diatomics in Molecules (DIM)

With the definitions made already it is now straightforward to write explicit expressions for the diatomics in molecules (DIM) local correlation model for the MP2 energy. Instead of retaining all significant doubles amplitudes, we shall instead retain only the direct product of ASIS functions, as specified by Equation (74). Therefore Equation (68) for the doubles amplitudes may be rewritten as follows in the ASIS basis, while Equation (82) for the MP2 energy is given immediately below:

$$\left[D_{KM}G_{LN} + G_{KM}D_{LN}\right]t^{MN} + \mathbb{I}_{KL} \tag{83}$$

$$E_{DIM-MP2} = \tfrac{1}{2}I_{KL}t^{KL} \tag{84}$$

We have introduced the ASIS metric matrix $\mathbf{G}$, and energy difference matrix $\mathbf{D}$ defined previously in Equations (77) and (78) respectively.

As was implicit in the previous section describing canonicalizing the ASIS basis, iterative solution of Equations (83) is a very challenging problem due to the very large condition number of the ASIS overlap matrix, $\mathbf{G}$. Hence, at present the only viable method for obtaining the DIM energy is to do it non-iteratively, by direct evaluation in the COASIS basis. The unsymmetrized and antisymmetrized integrals must be transformed to the COASIS basis, using the transformation defined by Equation (81), and the energy may then be obtained explicitly as:

$$E_{DIM-MP2} = -\frac{1}{2}\sum_{\bar{P}\bar{Q}}\frac{I_{\bar{P}\bar{Q}}\mathbb{I}_{\bar{P}\bar{Q}}}{E_{\bar{P}}E_{\bar{Q}}} \tag{85}$$

Equation (85) for the DIM-MP2 energy bears considerable similarity to the normal expression for the canonical MP2 energy, but as a result of performing manipulations in the atom-centered ASIS basis, we find:

(1) There are a quadratic number of energy contributions instead of the usual quartic number. We shall use the notation that capital letters denote extensive quantities such the AO basis size N, or the minimal basis size, M, while lower case letters denote intensive quantities, such as the number of minimal basis functions per atom, m, or the number of AO's per atom, n. The number of substitutions has been reduced from the usual O^2V^2 to m^2N^2 this way, which is a saving proportional to the number of atoms squared.

(2) The rate-determining computation steps will be cubic rather than 5^{th} order in the size of the molecule. Such steps will correspond to the COASIS canonicalization, and transformations of the unsymmetrized and antisymmetrized two-electron integrals into the COASIS basis, as required for Equation (85). These steps involve computational effort proportional to m^3N^3.

4.8 Triatomics in Molecules (TRIM)

The TRIM substitutions were described symbolically in Equation (75): this local correlation model applies the ASIS approximation to only one of the two pairs of single substitution indices that comprise the double substitutions. Unfortunately, the energy associated with the TRIM model cannot easily be obtained without iterations. The TRIM ansatz is not a direct product, and therefore the coefficient matrix coupling the retained substitutions in for example Equation (68) does not decompose into direct products of smaller matrices. However since the model still involves application of the ASIS approximation, we expect the condition number of the coefficient matrix to be extremely large (even after discarding any null space of zero eigenvalues). Therefore we are not optimistic about being able to efficiently converge such equations by iterative methods. The focus of our efforts has been to develop a viable noniterative alternative. The result of these investigations is described below.

4.8.1 Half-TRIM Energy

If we adopt a triatomics in molecules model that included only half of the substitutions that are incorporated in Equation (75), the resulting model has direct product structure:

$$\{aibj\}_{half-TRIM} = \{(ai)\} \otimes \{b\} \otimes \{j\} \tag{86}$$

As a result, it can be solved noniteratively by the same COASIS recanonicalization we developed earlier, and then applied above to develop an explicit expression for the DIM-MP2 energy.

The half-TRIM energy and amplitude equations may be written out in general form as restrictions of Equations (66) and (68), using the ASIS overlap and energy difference matrices **G** and **D**. The resulting equations may also be compared and contrasted with the corresponding DIM equations, (83) and (84):

$$\big[D_{KL}g_{bc}g_{jk} + G_{KL}\big(F_{bc}g_{jk} - g_{bc}F_{jk}\big)\big]t^{Lck} + \mathbb{I}_{Kbj} \tag{87}$$

$$E_{half-TRIM-MP2} = \tfrac{1}{2}I_{Kbj}t^{Kbj} \tag{88}$$

The noniterative solution to the half-TRIM MP2 problem employs the COASIS basis for the atomic single substitution indexes, thus bringing **G** and **D** to diagonal form. The remaining occupied and virtual index can be represented in the conventional canonical orthogonal basis (denoted also by overbars), so that the energy may be directly evaluated from the following expression:

$$E_{half-TRIM-MP2} = -\frac{1}{2}\sum_{\bar{P}\bar{b}\bar{j}} \frac{I_{\bar{P}\bar{b}\bar{j}} I\!I_{\bar{P}\bar{b}\bar{j}}}{E_{\bar{P}} + \epsilon_{\bar{b}} - \epsilon_{\bar{j}}} \tag{89}$$

The computational requirements of Equation (89) are roughly as follows:

(1) The number of substitutions is reduced from growing as the 4^{th} power of molecule size ($O^2 V^2$) to the 3^{rd} power of molecule size: specifically $mNOV$. This saves a factor proportional to the number of atoms.
(2) The significant computational steps scale either with the 3^{rd} and 4^{th} powers of the size of the molecule. Obtaining the transformation to the COASIS basis is m^3N^3, as for the DIM model, but the transformation of the two-electron integrals now involves a step proportional to the 4^{th} power of molecule size, in addition to steps scaling with the 3^{rd} power of molecule size. A factor at least proportional to the number of atoms can be saved relative to conventional transformations.

It is also worth noting that a compromise for the evaluation of the half-TRIM energy is possible, in which the ASIS indices are canonicalized (because the problem of large condition numbers has its origin in the ASIS approximation), but the remaining occupied and virtual indices are treated in a local basis. This may be the method of choice for the largest calculations using this method, because sparsity can then be exploited in the treatment of the nonlocal substitutions. Asymptotically the computational effort in this formulation will scale as the cube of molecule size, corresponding to the COASIS steps.

4.8.2 TRIM Energy Additivity Expression

The half-TRIM model will not be very close to the full TRIM model. We can predict that it will give an energy that is approximately half way between the DIM and TRIM models, because it includes only half of the additional substitutions in TRIM, while nevertheless using an unmodified energy expression. The energy expression, Equation (89), includes a factor of 1/2 based on double-counting all contributions, as is done in the DIM model. Yet the half-TRIM model includes only one copy of the new substitutions, and therefore their contribution to the energy will be undercounted by a factor of two. Recovering only about 50% of the TRIM-DIM correction is an unsatisfactory approximation, and we must do better.

The most obvious solution is to consider approximating the TRIM model by modifying the energy expression to correct such undercounting. We cannot simply double the half-TRIM energy, because then we over-count the DIM substitutions by a factor of two. However this can be corrected by subtracting the DIM energy itself, under the assumption that the omitted substitutions do not affect the calculated values. This result is the TRIM model we shall implement in practice. The associated energy is simply:

$$E_{TRIM-MP2} = 2E_{half-TRIM-MP2} - E_{DIM-MP2} \tag{90}$$

While Equation (90) is probably the simplest noniterative way to approximate the TRIM model, there are also other possibilities. However, based on numerical tests, we have not yet found any alternative that is superior to Equation (90).

4.9 Numerical Tests of the DIM And TRIM Models

Our primary purpose in this part of the chapter has been to discuss the development of the theory underlying the atoms in molecules models of local electron correlation. It is nonetheless still interesting to present a small selection of results that establishes the comparative performance of the local DIM and TRIM models relative to untruncated ("full") MP2 theory. The local MP2 methods have been implemented in a development version of the Q-Chem program suite[47] , which was used for all calculations. The results reported here are taken from our recent report on the TRIM model[23]. All calculations employ EPAO's to span the occupied space.

Details of the efficient implementation of the theory described above are reported in a recent thesis[48], and a paper in preparation. Here we simply summarize the capabilities of our program. The computational advantage associated with these local MP2 methods varies depending upon the size of molecule and the basis set. As a rough general estimate, TRIM-MP2 calculations are feasible on molecule sizes about twice as large as those for which conventional MP2 calculations are feasible on a given computer, and this is their primary advantage. Our implementation is well suited for large basis set calculations, because the memory requirement for the integral transformation does not exceed OON, and is thresholded so that it asymptotically grows linearly with molecule size. Additional memory of approximately $32N^2$ is required to complete the local MP2 energy evaluation. The disk space requirement is only about $8OVN$, but is not thresholded. The integrals are evaluated four times. DIM-MP2 calculations are faster than TRIM-MP2 and do not require disk storage, but have similar memory requirements.

4.9.1 Asymptotic Correlation Energy Recovery in Polyenes

The simplest measurement of accuracy for any local MP2 method is the percentage of untruncated MP2 correlation energy recovered. Any local method is exact for an atom, and many are exact for diatomic molecules. Therefore the quantity of interest is the limiting percentage of correlation energy recovered as molecule size becomes large. In Figure 1, the percentage of the full MP2 correlation energy recovered by the DIM and TRIM methods is plotted for each additional increment to a polyene chain respectively. The 6-31G* basis was used for these calculations.

First, let us examine the TRIM results, which are quite remarkable. It is clear that virtually no correlation energy is lost as a result of discarding what is asymptotically almost all of the double substitutions (the fraction kept is proportional to the inverse of the number of atoms). So little correlation energy is lost that the asymptotic correlation energy recovery is more than 5 times better than is commonly reported with the Saebø-Pulay model, which is perhaps typically 98%. This very high fidelity in recovering absolute correlation energy bodes well for the ability of the TRIM model to reliably reproduce relative energies also, as we investigate in the following two subsections. We also note that while the TRIM correlation energy

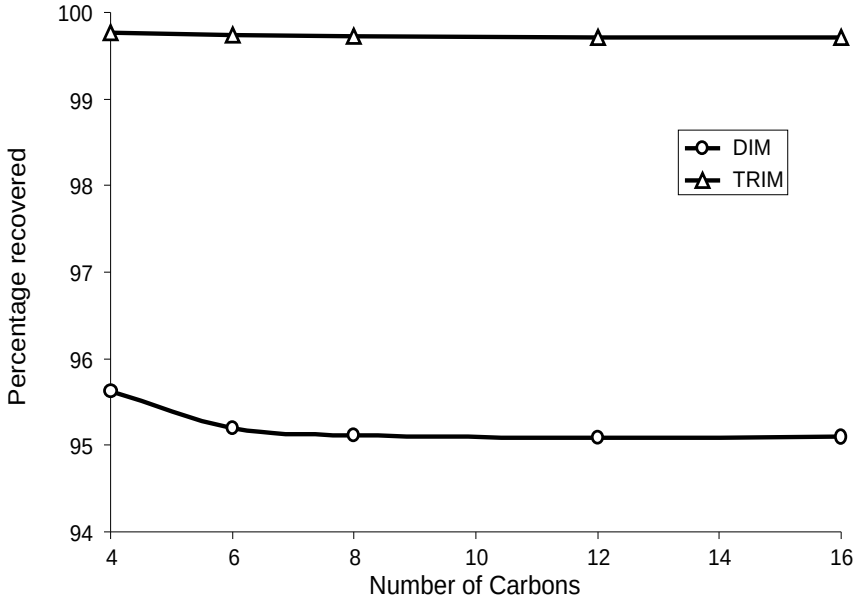

Figure 1. Percentage of the full (untruncated) MP2 correlation energy recovered by the diatomics in molecules (DIM) and triatomics in molecules (TRIM) local correlation methods for all-trans polyene chains of increasing length. Since local correlation models are usually exact for atoms and exhibit gradually decreasing fractional correlation energy recovery as the size of the molecule increases, we report the fractional correlation energy recovery for just the new increment of chain length. The 6-31G* basis was used for these calculations, and all orbitals were correlated.

recovery is always below 100% in the results reported here, it does not appear to obey a Hylleraas type variational principle with respect to the full MP2 correlation energy.

The second point of interest is the performance of the DIM method, which keeps a fraction of the double substitutions that is proportional to the inverse of the number of atoms squared. While no explicitly ionic substitutions are retained, the fraction of correlation energy recovered drops only by about 5% or so. This is about as good as could reasonably be expected, and suggests that the use of linearly dependent nonorthogonal functions must be accounting for a good fraction of local (nearest neighbor) ionic substitutions. Of course, one must also remember the large magnitude of the correlation energy (roughly 1 eV per spatially proximate pair of electrons) versus the small values of relative energies that are often of interest in chemistry. In this light it is nevertheless unclear if the DIM correlation energy recovery is sufficient to reliably treat relative energies.

4.9.2 Rotational Barriers in Conjugated Molecules

The second set of test calculations are MP2 corrections to the rotational barriers of vinyl and phenyl- substituted molecules[49]. This is a rather stringent test of the accuracy of a local MP2 method given that the barrier is a small difference in energy relative to the absolute energy. Thus it provides a reasonable assessment of the stability of the correlation energy recovery in a local method with respect to changes in the potential energy surface. Furthermore, the molecules are conjugated and the origin of the barriers is (nonlocal) resonance stabilization, which is also particularly challenging for local correlation methods. Thus success in this class of test problems bodes extremely well for success in the reproduction of relative energies in general.

Table 1 contains a summary of the individual results and the RMS and maximum errors (relative to untruncated MP2) as a result of (a) neglecting correlation, (b) employing DIM, and, (c) employing TRIM. The results indicate that TRIM deviates on average by only 0.03 kcal/mol, when compared to the full MP2 results. The DIM results show deviations nearly ten times as large (0.2 kcal/mol). The DIM deviations are nevertheless substantially smaller than simply neglecting electron correlation effects. Therefore the main conclusions are that the TRIM results reproduce untruncated MP2 to a precision which is more than satisfactory for routine chemical applications, and even DIM is certainly acceptable for this quite challenging set of problems.

It is also important to emphasize that both the TRIM and DIM results for relative energies are significantly better than the reproduction of total MP2 correlation energies might imply. In other words there is significant cancellation of errors in the evaluation of relative energies. The extent to which this is so can be partly answered from Table 1. Suppose the absolute MP2 correction to the energy of a vinyl-substituted molecule is 0.8 hartree or 500 kcal/mol (or about double, or 1000 kcal/mol for a substituted benzene). The TRIM method, for instance, might produce an error anywhere from 0.1% to 0.2% leading to a range of 0.5 kcal/mol. (or 1 kcal/mol for the substituted benzenes). Nonetheless, the TRIM rotational barriers are well within 0.1 kcal/mol. Thus the percentage of MP2 correlation energy recovered is quite stable with respect to conformational changes, due in large part to the associated smoothness of the EPAO description of the occupied space. This argument is even more important in accounting for the fairly reasonable results obtained with the DIM model.

4.9.3 Water Dimer Interaction Energy

The water dimer exhibits a significant correlation contribution to the hydrogen-bonding energy of binding. Furthermore due to the donor-acceptor nature of hydrogen bonding, it is possible that the correlation effects are not highly localized. The water dimer is thus a good system to assess the performance of the DIM and TRIM models for treating hydrogen-bonding interactions. MP2 energy corrections to the association energies of water clusters were assessed with the aug-cc-pVXZ series of basis sets, where X=D,T,Q. The geometry of the water dimer was obtained from a CCSD(T)/aug-cc-pVTZ calculation[50] . The results, for the correlation con-

Table 1. MP2/6-311G** corrections to Hartree-Fock relative energy differences in twelve conjugated molecules (kcal/mol)[a,b]. The conjugated molecules are substituted ethylenes of the form Vi-X, where Vi is the vinyl group, and substituted benzenes of the form Ph-X, where Ph is the phenyl group.

Molecule	**Conformation**	**TRIM**	**DIM**	**null**	**Full MP2**
Vi-C_2H_3	rotation	-0.249	0.030	0.000	-0.288
Vi-C_2H_3	gauche	-0.508	-0.182	0.000	-0.530
Vi-CFO	rotation	-0.913	-0.926	0.000	-0.930
Vi-CHO	rotation	-0.064	-0.028	0.000	-0.085
Vi-NH_2	rotation	0.770	0.605	0.000	0.806
Vi-NO_2	rotation	-1.795	-1.717	0.000	-1.834
Vi-OH	rotation	1.012	1.063	0.000	0.999
Vi-OH	anti	0.058	-0.084	0.000	0.059
Ph-OH	rotation	0.824	1.033	0.000	0.810
Ph-NO_2	rotation	-2.296	-2.524	0.000	-2.316
Ph-C_2H_3	rotation	-0.423	0.156	0.000	-0.506
Ph-CFO	rotation	-1.423	-1.694	0.000	-1.423
Ph-CHO	rotation	-0.459	-0.405	0.000	-0.493
Ph-NH_2	rotation	0.579	0.616	0.000	0.572
Ph-NH_2	inversion	0.735	0.548	0.000	0.741
RMS Error		0.031	0.252	1.022	
Max Error		0.083	0.662	2.316	

a All calculations were performed at HF/6-31G* optimized geometries, with all electrons correlated.

b The results exclude the Hartree-Fock contribution, values for which have been reported in ref. 49.

Table 2. Comparison of several local MP2 models for the calculation of the MP2 correction to the water dimer[a] association energy[b] (kcal/mol). These models are the triatomics in molecules (TRIM) model, as developed here, the Saebø-Pulay model (SP), and the diatomics in molecules (DIM) model.

Basis Set	**Full**	**TRIM**	**% Full**	**SP**[c]	**% Full**	**DIM**	**% Full**
aug-cc-pVDZ	-1.406	-1.316	94%	-0.52	37%	-0.191	14%
aug-cc-pVTZ	-1.541	-1.478	96%	-0.96	62%	-0.503	33%
aug-cc-pVQZ	-1.482	-1.432	97%	-1.17	79%	-0.804	54%

a All calculations were performed at the CCSD(T)/aug-cc-pVTZ optimized geometry[50].
b Only valence orbitals are correlated. The current best estimate[50,53] of the complete basis set limit frozen core MP2 value at the Halkier geometry[50] is 1.37 kcal/mol.
c Frozen core Saebø-Pulay local MP2 results from the calculations of Schütz *et al.*[54], also at the Halkier geometry.

tribution only, are reported in table 2. No counterpoise corrections were performed, because we wish to assess the convergence of the calculations towards the estimated complete basis set limit of 1.37 kcal/mol[50,51,52,53].

From table 2, the TRIM model closely tracks the untruncated MP2 values, with the difference gradually decreasing to 0.06 kcal/mol by the aug-cc-pVQZ basis. Table 2 also contains calculations using the SP model from ref. 54 and DIM calculations. The SP model yields results that are intermediate between the DIM and TRIM models, as might be anticipated from the discussion in Sec. 4.4.

It is noteworthy that the fraction of the correlation contribution to the binding energy that is recovered by the SP and DIM models increases strongly as the basis set size increases. This reflects an important point about these (and indeed all) local truncations. Both DIM and SP are exact in the limit in which each atom's set of basis functions approach global completeness. Therefore the quality of results obtained by these truncations improves markedly as the basis set size increases. The much milder truncation involved in the TRIM model results in less pronounced basis set dependence of the truncation effects. The SP model, and in particular the DIM model, recover a disturbingly low fraction of the interaction energy for the smaller basis sets, and therefore should probably not be used for hydrogen-bonding problems except with the larger basis sets.

5 An Overview of Other Applications

Space and time considerations prohibit us from giving an equally detailed discussion of our other recent applications of tensor ideas to electronic structure theory of many-electron systems. Instead, the following short discussion provides an intro-

ductory overview of several related research projects that we have been pursuing. Full details are available in the primary literature. The purpose of these short discussions is to give more examples of the usefulness of tensor methods in electronic structure theory.

5.1 Local Electron Correlation Models for Triple Substitutions

The local correlation methods discussed above for double substitutions, in the context of MP2 theory, can be extended to yield local correlation methods for triple substitutions[55,56]. Triple substitutions are essential for predicting chemical reaction energies to 1 kcal/mol or better, and are employed in the widely used G2 and G3 thermochemical methods, where the rate-determining step is usually the triples contribution to the fourth order Møller-Plesset energy: MP4(T). The cost of methods including triple substitutions, such as the widely used MP4(T) and CCSD(T) methods, scales as the seventh power of molecule size, and this limits their application to roughly ten first row atoms.

The basis of the local triples method is a generalization of the TRIM method discussed above for MP2 theory. The same nonorthogonal sets of functions are used to span the occupied space (the molecule-adapted minimal basis, projected into the occupied space), and the virtual space (the full atomic orbital basis, projected into the virtual space). Three electrons are simultaneously promoted in a triple substitution, and in the local model we have developed, two of those promotions are restricted to a single atom, while the third is permitted to be non-local. This reduces the number of triple substitutions from rising with the 6^{th} power of molecule size to rising with the 4^{th} power. The recanonicalization described in the section on the TRIM model for MP2 theory is even more valuable in the context of triple substitutions, because it means that the local triples amplitudes do not need to be stored. Instead they can be made in the recanonicalized representation in batches, their contribution to the triples energy can be immediately obtained, and they can be discarded prior to commencing the next batch.

Benchmark thermochemical calculations on 105 molecules in the G2/97 database indicate that local truncation recovers at least 95% of the untruncated triples energy. The local error introduced into the G2 binding energies is typically 0.1 - 0.2 kcal/mol, with a maximum error of 0.26 kcal/mol. This error is small enough for most applications. Finally, while the efficient computational implementation of this local triples method is a challenging algorithmic problem, the results are worthwhile[56]. The local triples algorithm reduces the computational cost from growing with the seventh power of molecule size to the fifth power of molecule size. In practice, this enables triples calculations on molecules two to three times larger than previously feasible. It crosses over with the conventional triples algorithm around 25 occupied orbitals, or roughly 150 basis functions. The local algorithm requires all the doubles amplitudes to be stored on disk, and also requires disk storage for a cubic number of local integrals. A relatively small (quadratic) amount of memory is required.

The discussion of atoms in molecules models of local correlation in Section 4.3 listed several different atomic truncations of the full set of double substitutions. The gentlest truncation is triatomics in molecules, where one occupied to unoccupied orbital substitution must occur on a single atom, so that the overall double substitution cannot involve more than 3 atoms. In the more strongly truncated diatomics in molecules model, two orbital substitutions are each required to be on single atoms, so that the double substitution cannot involve more than a pair of atoms. The most drastic truncation, which we termed atoms in molecules, requires all 4 orbitals involved in the double substitution to be on a single atom. Such an approximation would not correctly describe the dispersion energy, and therefore we did not consider it as a useful approximation in the context of MP2 theory.

Are there other contexts in which the very drastic atoms in molecules approximation to electron correlation (or something very similar to it) could be useful? The answer is yes, and in this section we shall describe one such context: namely the so-called perfect pairing approaches to electron correlation[57,58]. In perfect pairing, the (unnormalized) wavefunction is taken to be an antisymmetrized (A) product of doubly occupied "core" functions, and geminal (2-electron) functions, g_i that describe electron correlations within each valence bonded pair (and each lone pair) of electrons:

$$\big|\Psi\big\rangle = A\big[\varphi_1\bar{\varphi}_1\varphi_2\bar{\varphi}_2\cdots g_1 g_2\cdots\big] \tag{91}$$

$$g_i = A\big[\varphi_{C+i}\bar{\varphi}_{C+i} + t_i\varphi^*_{C+i}\bar{\varphi}^*_{C+i}\big] \tag{92}$$

Here C is the number of doubly occupied (C)ore orbitals. Clearly a wavefunction of this form is well suited to breaking single bonds, where at dissociation, the bonding and antibonding orbitals, φ_i and φ^*_i, approach degeneracy.

It has been known for some time[59], although it is still not widely appreciated, that the PP wavefunction can be exactly rewritten in the form of a simplified coupled cluster doubles wavefunction; no approximation is involved. Specifically:

$$\big|\Psi\big\rangle = exp\big(T_2^{PP}\big)\big|\Phi_0\big\rangle \tag{93}$$

$$\big|\Phi_0\big\rangle = A\big[\varphi_1\bar{\varphi}_1\varphi_2\bar{\varphi}_2\cdots\varphi_O\bar{\varphi}_O\big] \tag{94}$$

$$\hat{T}_2^{PP}\big|\Phi_0\big\rangle = \sum_i t_i\big|\Phi_i\big\rangle \tag{95}$$

The double substitution $|\Phi_i\rangle$ is the determinant $|\Phi_0\rangle$ with the occupied pair $\varphi_{C+i}\bar{\varphi}_{C+i}$ replaced by the unoccupied (correlating) pair $\varphi^*_{C+i}\bar{\varphi}^*_{C+i}$. Viewed in this light, the PP wavefunction is evidently a form of local correlation! It is a form in which each occupied pair is treated separately (the pair is the atomic analog here).

Conventionally the bonding and antibonding functions that comprise the wavefunction are taken to be strongly orthogonal to each other. With this constraint the energy can be solved for either variationally, or nonvariationally via the machinery of coupled cluster theory. The resulting method is qualitatively quite successful in describing potential energy curves. Single, double, and even triple bond-breaking curves are qualitatively correct using perfect pairing. This applies whether the energy is obtained via the variational method, or via the coupled cluster approach, which is not formally variational.

At the same time, PP does have some notable deficiencies, which are directly connected to the limited description of electron correlation that it contains. For example, applying PP to the benzene molecule leads to symmetry breaking: PP predicts that the most stable geometry has alternating carbon-carbon bond lengths rather than equal ones. This arises from the fact that there are two equivalent sets of localized occupied orbitals that can describe the π electrons, which can be loosely mapped to the two Kekule structures. When one of these sets is chosen, and a corresponding set of correlating orbitals are defined, this favors localization of the electronic structure. The extent of symmetry breaking is relatively small in energy terms[60]: about 3 kcal/mol is the energy difference between D_{6h} and D_{3h} structures. Nevertheless this is clearly an undesirable result.

Given our use of nonorthogonal functions to develop local correlation models for double substitutions, we thought it natural to ask whether nonorthogonal functions might be useful for alleviating the problem discussed above for the PP model. In particular, we have explored the possibility of using the following redefinition of the orbitals in Equation (92). Instead of permitting one orthogonal occupied and one correlating orbital for each valence pair of *electrons*, we prefer to use one occupied and one correlating orbital for each valence atomic *orbital*[60]. Since the number of valence atomic orbitals is larger than the number of electrons, this is clearly a redundant, or linearly dependent representation. The occupied valence orbitals will be nonorthogonal to each other, and the unoccupied correlating orbitals will also be nonorthogonal to each other. The two sets will remain strongly orthogonal, however, as is necessary to make the occupied space well defined.

The use of redundant representations for the *one-particle* (orbital) space permits additional flexibility in the *two-particle* space, because there are now more amplitudes, and these additional amplitudes are not entirely redundant. In the case of benzene in the π space, there will now be one occupied orbital for each C atom (corresponding essentially to a p_z function projected into the occupied space), and one correlating orbital. Clearly, each symmetry-equivalent atom is treated equivalently as far as correlation is concerned, and so it is not surprising that the resulting nonorthogonal PP model[60] (solved via coupled cluster methods), eliminates the artifactual symmetry breaking seen in the traditional PP model applied to benzene.

5.3 Second Order Corrections to Singles and Doubles Coupled Cluster Energies

To approach chemical accuracy in reaction energies and related properties, it is necessary to account for electron correlation effects that involve 3 electrons simul-

taneously, as represented by triple substitutions relative to the mean field single determinant reference, which arise in MP4. The best standard methods for including triple substitutions are the CCSD(T) type methods[61]. The accuracy of these methods is well-documented for many cases[62] and in general is a very significant improvement relative to the starting point. While the (T) corrections have been extraordinarily successful, there is nonetheless still room for improvement. They contain judiciously chosen terms from 4^{th} and 5^{th} order Møller-Plesset perturbation theory, as well as higher order terms that result from the fact that the converged cluster amplitudes are employed to evaluate the 4^{th} and 5^{th} order terms. The correction therefore depends upon the bare reference orbitals and orbital energies, and in this way its effectiveness still depends on the quality of the reference determinant. Since we are correcting a coupled cluster solution rather than a single determinant, this is an aspect of the (T) corrections that can be improved.

Such an improvement has recently been reported[63]. The new correction is a true second order correction to a coupled cluster starting point and is therefore denoted as (2). It was first developed for the coupled cluster doubles method using optimized orbitals[64] (denoted as OD) and is also straightforward to implement for either CCSD or QCISD, given the general theory already presented[65]. The basis of the (2) method is to partition not the regular Hamiltonian into perturbed and unperturbed parts, but rather to partition a similarity-transformed Hamiltonian, defined as $\hat{\bar{H}} = e^{-\hat{T}} \hat{H} e^{\hat{T}}$. In the truncated space (call it the p-space) within which the cluster problem is solved (e.g. singles and doubles for CCSD), the coupled cluster wavefunction is a true eigenvalue of $\hat{\bar{H}}$. Therefore we take the zero order Hamiltonian, $\hat{\bar{H}}^{(0)}$, to be the full $\hat{\bar{H}}$ in the p-space, while in the space of excluded substitutions (the q-space) we take only the one-body part of $\hat{\bar{H}}$ (which can be made diagonal). The fluctuation potential describing electron correlations in the q-space is $\hat{\bar{H}} - \hat{\bar{H}}^{(0)}$, and the (2) correction then follows from second order perturbation theory.

Tensor methods are naturally useful in the development of the detailed expressions for the (2) correction[65]. This is because the similarity transformed Hamiltonian, $\hat{\bar{H}}$, is non-Hermitian. As a result, when the one-body part of $\hat{\bar{H}}$ is made diagonal, the single particle eigenfunctions that result are different on the left and right sides. In other words, the single particle bras and kets are different in the basis in which we work. Furthermore, in this semicanonical basis, the bras and kets are biorthogonal to each other, but each set contains functions that are nonorthogonal amongst themselves. This basis defines the natural representation that is appropriate for this problem, and in ref. 65, the working equations of the (2) methods are derived and presented in this representation, as well as more general biorthogonal natural representations (in which the single particle $\hat{\bar{H}}$ is not diagonal).

The new partitioning of terms between the perturbed and unperturbed Hamiltonians inherent in the (2) correction leads to a correction that show both similarities and differences relative to the existing (T) corrections. There are two types of higher correlations that enter at second order: not only triple substitutions, but also quadruple substitutions. The quadruples are treated with a factorization ansatz, that is exact in 5^{th} order Møller-Plesset theory[66] , to reduce their computational

cost from N^9 to N^6. For large basis sets this can still be larger than the cost of the triples terms, which scale as the 7^{th} power of molecule size, with a factor twice as large as the usual (T) corrections. These corrections are feasible for molecules containing between four and ten first row atoms, depending on computer resources, and the size of the basis set chosen. There is early evidence that the (2) corrections are superior to the (T) corrections for highly correlated systems[63]. This shows up in improved potential curves, particularly at long range. For such problems, with restricted orbitals, the (T) corrections tend to diverge, but the new (2) corrections are stable provided the coupled cluster reference itself is stable. This advantage may also extend to improved energetic and structural properties at equilibrium in problematical cases. It will be some time before sufficient testing on the new (2) corrections has been done to permit a general assessment of the performance of these methods, but they are clearly very promising.

6 Conclusions

In this chapter we have tried to accomplish several objectives. First, we have provided a general introduction to tensor methods in the context of electronic structure theory, where they are particularly well-suited to dealing with nonorthogonal single particle basis sets. Second, we have discussed local electron correlation models for second order Møller-Plesset theory, that are based entirely on nonorthogonal single particle basis sets both as an application of the tensor algebra and as an illustration of the potential usefulness of nonorthogonal functions. Third, we have provided a brief description of several other applications in electronic structure theory where we have found the tensor methods useful, to further illustrate their value. Fourth, we would like to emphasize that there are many other potential applications awaiting us in the future. Perhaps this is the best reason to at least be generally aware of how to apply tensor methods in electronic structure theory.

Acknowledgments

This work was supported by grants from the National Science Foundation, the Air Force Office of Scientific Research, and the Department of Energy. MHG gratefully acknowledges a Packard Fellowship (1995-2000).

References

1. For an introductory treatment of electron correlation methods, see for example, A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry* (McGraw-Hill, New-York, 1989).
2. R. McWeeny, *Methods of Molecular Quantum Mechanics* (Academic Press, London, 1989).
3. F. E. Harris, H. J. Monkhorst, and D. L. Freeman, *Algebraic and Diagrammatic Methods in Many-Fermion Theory* (Oxford University Press, Oxford, 1992).
4. E. B. Stechel, T. A. Schmalz, and J. C. Light, J. Chem. Phys. **70**, 5640 (1979).
5. O. Sinanoglu, Theor. Chim. Acta **65**, 223 (1984).

6. E. Artacho and L. Milans del Bosch, Phys. Rev. A **43**, 5770 (1991).
7. M. Head-Gordon, P. E. Maslen and C. A. White, J. Chem. Phys. **108**, 616 (1998).
8. For examples of formation of the effective Hamiltonian with linear scaling computational effort, see, C. A. White, B. G. Johnson, P. M. W. Gill and M. Head-Gordon, Chem. Phys. Lett. **253**, 268 (1996), and references therein.
9. For a discussion of linear scaling methods that replace diagonalizing the Hamiltonian, see, R. Baer and M. Head-Gordon, Phys. Rev. Lett. **79**, 3962 (1997), and references therein.
10. P. Pulay and S. Saeb, Ann. Rev. Phys. Chem. **44**, 213 (1993).
11. P. Fulde, *Electron Correlations in Molecules and Solids*, 3rd edition (Springer-Verlag, Berlin, 1995).
12. J. L. Synge and A. Schild, *Tensor Calculus* (Dover, New York, 1978).
13. A. D. Myskis, *Advanced Mathematics for Engineers* (Mir, Moscow, 1979) Chapter 5.
14. D. Lovelock and H. Rund, *Tensors, Differential Forms and Variational Principles* (Dover, New York, 1988).
15. T. Frankel, *The Geometry of Physics: An Introduction* (Cambridge, Cambridge, 1997).
16. A. Edelman, T. A. Arias, and S. T. Smith, SIAM J. Matrix Anal. Appl. **20**, 303 (1998).
17. W. Kohn, Phys. Rev. **115**, 809 (1959).
18. S. L. Campbell and C. D. Meyer, Junior. *Generalized Inverses of Linear Transformations*, (Dover Publications, New York, 1991).
19. See for example, J. F. Stanton and R. J. Bartlett, J. Chem. Phys.**98**, 7029 (1993), and references therein.
20. See for example, J. J. W. McDouall, Theor. Chim. Acta **83**, 339 (1992); **85**, 395 (1993), and references therein.
21. P. E. Maslen and M. Head-Gordon, Chem. Phys. Lett. **283**, 102 (1998).
22. P. E. Maslen and M. Head-Gordon, J. Chem. Phys. **109**, 7093 (1998).
23. M. S. Lee, P. E. Maslen and M. Head-Gordon, J. Chem. Phys. **112**, 3592 (2000).
24. C. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).
25. W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory* (Wiley, New York 1986).
26. T. Helgaker, J. Gauss, P. Jorgensen, and J. Olsen, J. Chem. Phys. **106**, 6430 (1997).
27. W. Kohn, A. D. Becke and R. G. Parr, J. Phys. Chem. **100**, 12974 (1996).
28. For the most recent developments in semidirect methods, together with a review of earlier work, see, M. Head-Gordon, Mol. Phys. **96**, 673 (1999).
29. P. Pulay, Chem. Phys. Lett. **100**, 151 (1983).
30. S. Saebo, P. Pulay, J. Chem. Phys. **86**, 914 (1987).
31. R. B. Murphy, M. B. Beachy, R. A. Friesner, and M. N. Ringnalda, J. Chem. Phys. **103**, 1481 (1995).
32. G. Hetzer, P. Pulay, and H. J. Werner, Chem. Phys. Lett. **290**, 143 (1998).
33. G. Rauhut, P. Pulay, and H. J. Werner, J. Comput. Chem. **11**, 1241 (1998).

34. M. Schütz, G. Hetzer, and H. J. Werner, J. Chem. Phys. **111**, 5691 (1999).
35. C. Hampel and H. J. Werner, J. Chem. Phys. **104**, 6286 (1996).
36. G. Reynolds, T. J. Martinez, and E. A. Carter, J. Chem. Phys. **105**, 6455 (1996).
37. G. Reynolds and E. A. Carter, Chem. Phys. Lett. **265**, 660 (1997).
38. J. W. Boughton and P. Pulay, J. Comput. Chem. **14**, 736 (1998).
39. A. El Azhary, G. Rauhut, P. Pulay and H.-J. Werner, J. Chem. Phys. **108**, 5185 (1998).
40. T. Nakajima and H. Nakatsuji, Chem. Phys. **242**, 177 (1999).
41. R. A. Friesner, R. B. Murphy, M. D. Beachy, M. N. Ringnalda, W. T. Pollard, B. D. Dunietz, Y. X. Cao, J. Phys. Chem. **A103**, 1913 (1999).
42. For one example, see, W. J. Lauderdale, J. F. Stanton, J. Gauss, J. D. Watts, and R. J. Bartlett, Chem. Phys. Lett. **187**, 21 (1991).
43. M. S. Lee and M. Head-Gordon, Int. J. Quantum Chem. **76**, 169 (2000).
44. M. S. Lee and M. Head-Gordon, J. Chem. Phys. **107**, 9085 (1997).
45. M. S. Lee and M. Head-Gordon, Comp. Chem. **24**, 295 (2000).
46. M. S. Lee and M. Head-Gordon (work in progress).
47. C. A. White, J. Kong, D. R. Maurice, T. R. Adams, J. Baker, M. Challacombe, E. Schwegler, J. P. Dombroski, C. Ochsenfeld, M. Oumi, T. R. Furlani, J. Florian, R. D. Adamson, N. Nair, A. M. Lee, N. Ishikawa, R. L. Graham, A. Warshel, B. G. Johnson, P. M. W. Gill and M. Head-Gordon, Q-Chem, Version 1.2, Q-Chem, Inc., Pittsburgh, PA (1998).
48. M. S. Lee, Ph. D. Thesis, University of California, Berkeley, 2000.
49. M. Head-Gordon and J. A. Pople, J. Phys. Chem. **97**, 1147 (1993).
50. A. Halkier, H. Koch, P. Jorgensen, O. Christiansen, I. M. Beck Nielsen and T. Helgaker, Theor. Chem. Acc. **97**, 150 (1997).
51. S. S. Xantheas, J. Chem. Phys. **104**, 8821 (1996).
52. M. Schütz, S. Brdarski, P. O. Widmark, R. Lindh, and G. Karstrom, J. Chem. Phys. **107**, 4597 (1997).
53. W. Klopper, and H. P. Lüthi, Mol. Phys. **96**, 559 (1999).
54. M. Schütz, G. Rauhut, and H.-J. Werner, J. Phys. Chem. **A102**, 5997 (1998).
55. P. E. Maslen, M. S. Lee and M. Head-Gordon, Chem. Phys. Lett. **319**, 205 (2000).
56. P. E. Maslen, M. S. Lee and M. Head-Gordon (in preparation).
57. A. C. Hurley, J. Lennard-Jones and J. A. Pople, Proc. Roy. Soc. London, Ser. A **220**, 446 (1953).
58. W. A. Goddard III, and L. B. Harding, Ann. Rev. Phys. Chem. **29**, 363 (1978).
59. J. Cullen, Chem. Phys. Lett. **202**, 217 (1996).
60. T. Van Voorhis and M. Head-Gordon, J. Chem. Phys. **112**, 5633 (2000).
61. K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).
62. See for example, T. J. Lee and G. E. Scuseria, in Quantum Mechanical Calculations with Chemical Accuracy, edited by S. R. Langhoff (Kluwer, Dordrecht, 1995) p 47.
63. S. R. Gwaltney and M. Head-Gordon, Chem. Phys. Lett. **323**, 21 (2000).

64. C. D. Sherrill, A. I. Krylov, E. F. C. Byrd, and M. Head-Gordon, J. Chem. Phys. **109**, 4171 (1998).
65. S. R. Gwaltney, C. D. Sherrill, M. Head-Gordon, and A. I. Krylov, J. Chem. Phys. **113**, 3548 (2000).
66. S. A. Kucharski and R. J. Bartlett, J. Chem. Phys. **108**, 5243 (1998); 108, 9221 (1998).